

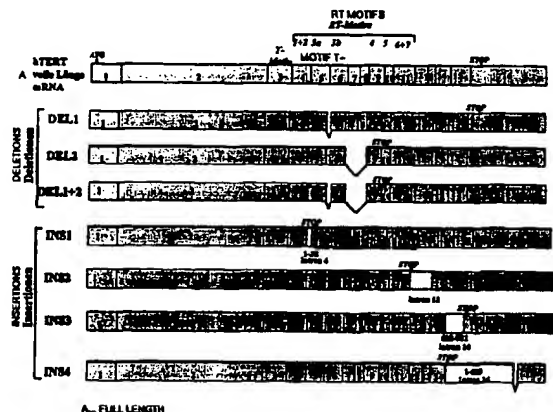
PCT
WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro
INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)



<p>(51) Internationale Patentklassifikation ⁶ : C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027</p>	A2	<p>(11) Internationale Veröffentlichungsnummer: WO 99/33998</p> <p>(43) Internationales Veröffentlichungsdatum: 8. Juli 1999 (08.07.99)</p>
<p>(21) Internationales Aktenzeichen: PCT/EP98/08216</p> <p>(22) Internationales Anmeldedatum: 22. Dezember 1998 (22.12.98)</p> <p>(30) Prioritätsdaten: 197 57 984.1 24. Dezember 1997 (24.12.97) DE</p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE).</p> <p>(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): HAGEN, Gustav [DE/DE]; Bertha-von-Suttner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggen-dorfstrasse 59, D-51061 Köln (DE).</p> <p>(74) Gemeinsamer Vertreter: BAYER AKTIENGE-SELLSCHAFT; D-51368 Leverkusen (DE).</p>	<p>(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Veröffentlicht <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p>	

(54) Title: **REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF**

(54) Bezeichnung: **REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG**



(57) Abstract

The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.

(57) Zusammenfassung

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen
Telomerase-Untereinheit und deren diagnostische und therapeutische Verwen-
dung

5 Aufbau und Funktion der Chromosomenenden

Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten
10 Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).

15 Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.

20 Desweiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).

25 Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der
30 Okazaki-Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-Strang das 5'-Ende, denn dort kann der RNA-Primer nicht durch DNA ersetzt

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

10

15 Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

Das Enzym Telomerase synthetisiert die Telomere

20

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

25

Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet werden, die als Multiproteinkomplex im Zellkern lokalisiert ist. Während der RNA-Anteil der humanen Telomerase schon seit längerem bekannt ist (Feng *et al.*, 1995),

30

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomeraseaktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34⁺38⁺-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumorgewebe eine signifikante Telomerase-Aktivität nachgewiesen (1734/2031, 85 %; Shay, 1997), während in normalem somatischem Gewebe keine Aktivität gefunden wurde (1/196, <1 %, Shay, 1997). Verschiedene Untersuchungen zeigten außerdem, daß in seneszenten Zellen, die mit viralen Oncoproteinen transformiert wurden, die Telomere weiterhin schrumpften und Telomerase nur in der Subpopulation entdeckt werden konnte, die die Wachstumskrise überlebte (Counter *et al.*, 1992). In diesen immortalisierten Zellen waren auch die Telomere stabil (Counter *et al.*, 1992). Ähnli-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgenese ist.

5 Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomersequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase
10 enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein. Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomersequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus
15 von Tumorzellen zur Stabilisation ihrer Telomere.

Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzweiligen Strahlen schädigen nicht nur die Tumorzellen,
20 sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle
25 Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden
30 dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapie von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und post-translationale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein1“) spezifisch gebunden werden.

20

25

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements“) be-

30

schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebes- oder zellspezifische Promotoreigenschaften.

Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen katalytischen Telomerase Untereinheit gemäß Fig. 10 (SEQ ID NO 3).

Die Erfindung betrifft weiterhin regulatorisch wirksame Teilbereiche der 5'-flankierenden regulatorischen DNA-Sequenz gemäß Fig. 4 (SEQ ID NO 1).

Weiterhin sind Gegenstand der vorliegenden Erfindung Intron-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit, insbesondere solche, die regulatorische Wirkung haben. Die erfindungsgemäßen Intronsequenzen werden im Rahmen von Beispiel 5 detailliert beschrieben (vgl. SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und 20).

Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon
5 beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere
10 weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15 Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).
20

Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25 Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- α , IFN- β , IFN- γ , TNF, TNF- α , TNF- β , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls
30 zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 α , - β), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF, M-CSF.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispielsweise:

Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle β -Glukuronidase, pflanzliche β -Glukuronidase aus *Secale cereale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle Carboxypeptidase, bakterielle β -Lactamase, bakterielle Cytosindeaminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC), β -Galaktosidase (β -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH), β -Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch
5 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch
10 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen,
15 zen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccina Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.
20

Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfindungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.
25

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

Die Zubereitungen der erfindungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-
30

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

5 Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilsequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF β , Insulin oder Insulin-like Growth Factor (ILGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intra-venös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, trans-dermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die syste-
20 mische Applikation dieser Zubereitungen.

Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen,
30 die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen
5 Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,

B. Messung des Substanzeffektes auf die Reportergenexpression.

10 Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.

15 Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.

20 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.

25 Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.

Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

30

- 5 A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reportergen mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten;
- 10 C. Vergleich des diagnostischen Werts mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

15 Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

20 Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

25 A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Sprague Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

30 B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten λ -DNA des Phagenklons P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene λ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des
10 rekombinanten Phagens. Die Pfeile kennzeichnen die zwei λ -Arme des Vektors EMBL3 Sp6/T7.

Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten λ -DNA des Phagenklons, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

15 A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene λ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur
20 14: nicht verdaut.

B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

30

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

10

15

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 5'GTTAAGTTGTAGCTTACACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

20

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die dargestellte Sequenz entspricht SEQ ID NO 2.

25

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

30

Exons sind als durchnummerierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymchnittstellen sind wie folgt gekennzeichnet: S, SacI; X, XhoI. Die relative Anordnung der fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

10

15

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

20

25

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgespligte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

30

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.

5 11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

10

Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.

Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstrukte dargestellt. Das promotorlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite ist die relative Luziferaseaktivität der verschiedenen Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

15

20

25 Tab. 1: Exon Intron Übergänge des hTC-Gens

Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5' Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Großbuchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

30

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in
10 Intron 2.

Tab. 1

3' Acceptor Sequence			5' Donor Sequence					
Intron	Exon	Exon No.	bp	Exon	Intron	Intron on	bp	
						No.		No.
5' flanking Region								
cagggcgcttccccgag	GTTCAGGCAGCGCTGCGT	1	281	CGCCCCCTCCTTCGCCAG	gtgggctccccgggtcg	1	104	
catgtccttctctgttttag	GTGTCCTGCCTGAAGGAGC	2	1354	TGGCTGCGCAGGAGCCAG	gtgaggggtgtgtggcgt	2	8616	
gaggggctctctctattgag	GGGTGGCTGTGTTCGGGC	3	196	TGCAAAAGCATTTGGAATCAG	gtactgtatccccacgcca	3	2089	
cccatgctgtccccgcag	ACAGCACTTGAAGAGGGTG	4	181	GTTCCGCAGAGAAAAGAGG	gtggtgtgtgtttttag	4	687	
ctcgccctccactcacagag	GCCGAGCGTCTCACCTCGA	5	180	TGAGCTGTACTTTGTCAAG	gtgggtgccccgggaccccc	5	494	
ccctctctctctgcccggag	GTGGATGTGACGGGGCGGT	6	156	CAAGGCCTTCAAGAGCCAC	gttaggttcaagtgtgata	6	>4660	∞
ctcccgtgtgttttcgag	GTCTCTACCTTGACAGACC	7	96	TGCCGTGTCATCGAGCAG	gtctgggcaactgcccctgca	7	980	
ctgtgtcttccccgccag	AGCTCCTCCCTGAATGAGG	8	86	CCGTGCGCATCAGGGGCAA	gtgagtcaggttgccaggt	8	2485	
gtattttcccttatttttag	GTCTACGTCCAGTGCCAG	9	114	CGGGATTCCGGCGGACGG	gtgaggcctcctctctccccc	9	1984	
cattgccccctctgcttag	GCTGCTCCTGCGTTGGTG	10	72	ACGGGAAAACCTTCTCTCAG	gtgagccccgtgccgtgtg	10	1871	
attcccccatgtgtcttag	GACCTGGTCCGAGGTGTC	11	189	TGCAGAGCGACTACTCCAG	gtgagcgcaactggcccgga	11	3800	
tccttcttggcgactcttag	CTATGCCCGGACCTCCATC	12	127	CCTGTTCTGGATTTGCAG	gtgagcaggctgagtgtca	12	880	
ctgtccgcacatctcttag	GTGAACAGCCTCCAGACGG	13	62	TCTGTGTGACGGCGTACAG	gtgagccgccaccaagggg	13	3187	
agcctctgttttccccag	GTTTACGCATGTGTCTG	14	125	CTGAAAGCCAGAACGCGAG	gtatgtgcagggtgcctggc	14	781	
tcgtattttggccccgag	GGATGTGCTGGGGGCCAA	15	138	CTGGGTCACTCAGGACAG	gcaagtgtgtgtggagccc	15	536	
	CCCAGACGCAGCTGAGTCG	16	664	TTTTTCAGTTTTTGAATAAA	3' flanking Region			

Tab. 2

Faktoren	Lokalisation in Intron 2
C/EBP	2925
CRE.2	2749
Sp1	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
API	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen bzw. eventuell auch ghTC-Pseudogene existieren.

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschrte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

5 Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca $1,5 \times 10^6$ Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 µm; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa
10 500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 µg/ml
15 Heringssperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit $1,5 \times 10^6$ cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte
20 durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ-DNA Präparation dieses Phagens Ausubel *et al.*
25 (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen
30 Experimenten jeweils 1 bis $1,5 \times 10^6$ Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach λ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umklontiert (Beispiel 4).

Beispiel 3

Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde λ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).

Da die isolierte λ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

Beispiel 4

Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.

Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umkloniert. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5' Sequenz.

Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-Fragment sowie ein 1,8 kb großes BamHI-Fragment subkloniert. Der Kombinations-Restriktionsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.

Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subkloniert.

Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subkloniert. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenklons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenklon P17 und 5' von Phagenklon P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

10

15

Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-ACTATAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen kloniert.

20

25

30

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTTGAAGGCCTTG-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATAACC-3') zusammen mit der HDL-Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten
5 Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking
10 Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-GGTGGATGTGACGGGCGCGTACG-3') und des „nested“ genspezifischen Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen genomischen hTC-Region identifiziert (s. Fig. 7).

15 Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten
20 Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5 zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch
25 dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2 unterstrichen.

Contig1:

	ACTTGAGCCC	AAGAGTTCAA	GGCTACGGTG	AGCCATGATT	GCAACACCAC	ACGCCAGCCT	TGGTGACAGA	70
5	ATGAGACCCT	GTCTCAAAAA	AAAAAAGAAA	AATTGAAATA	ATATAAAGCA	TCTTCTCTGG	CCACAGTGGG	140
	ACAAAACCAG	AAATCAACAA	CAAGAGGAAT	TTTGAAGAACT	ATACAAACAC	ATGAAAAATTA	AACAAATATAC	210
	TTCTGAATGA	CCAGTGAGTC	AATGAAGAAA	TTAAAAAGGA	AATTGAAAAA	TTTATTTAAG	CAAATGATAA	280
	CGGAACATA	ACCTCTCAAA	ACCCACGGTA	TACAGCAAAA	GCAGTGCTAA	GAAGGAAGTT	TATAGCTATA	350
	AGCAGCTACA	TCAAAAAAGT	AGAAAAGCCA	GGCGCAGTGG	CTCATGCCTG	TAATCCACAG	ACTTTGGGAG	420
10	GCCAAGGCCG	GCAGATCGCC	TGAGGTCAGG	AGTTCGAGAC	CAGCCTGACC	AACACAGAGA	AACCTTGTCG	490
	CTACTAAAAA	TACAAAAATTA	GCTGGGCATG	GTGGCACATG	CCTGTAATCC	CAGCTACTCG	GGAGGCTGAG	560
	GCAGGATAAC	CGCTTGAAAC	CAGGAGGTGG	AGGTTCGGGT	GAGCCGGGAT	TGCGCCATTG	GACTCCAGCC	630
	TGGGTAACAA	GAGTGAAACC	CTGTCTCAAG	AAAAAAGAAA	AAGTAGAAAA	ACTTAAAAAT	ACAACCTAAT	700
	GATGCACCTT	AAAGAACTAG	AAAAGCAAGA	GCAAACTAAA	CCTAAAAATTG	GTAAAAAGAA	AGAAATAATA	770
15	AAGATCAGAG	CAGAAATAAA	TGAAACTGAA	AGATAACAAT	ACAAAAGATC	AACAAAAATTA	AAAGTTGGTT	840
	TTTTGAAAAG	ATAAACAAAA	TTGACAAACC	TTTGCCAGGA	CTAAGAAAAA	AGGAAGAAG	ACCTAAATAA	910
	ATAAAGTCAG	AGATGAAAAA	AGAGACATTA	CAACTGATAC	CACAGAAATT	CAAAAGGATCA	CTAGAGGCTA	980
	CTATGAGCAA	CTGTACACTA	ATAAATTGAA	AAACCTAGAA	AAAATAGATA	AATTCTTAGA	TGCTACAAAC	1050
	CTACCAAGAT	TGAACCTAGA	AGAAATCCAA	AGCCCAACAA	GACCAATAAC	AATAATGGGA	TAAAGGCCAT	1120
20	AATAAAAAAGT	CTCCTAGCAA	AGAGAAGCCC	AGGACCCAAT	GGCTTCCCTG	CTGGATTTTA	CCAATCATTT	1190
	AAAGAAAGAT	GAATTCCAAT	CCTACTCAAA	CTATTCTGAA	AAATAGAGGA	AAGAATACCT	CCAACCTCAT	1260
	TCTACATGGC	CAGTATTACC	CTGATTCCAA	AACCAGACAA	AAACACATCA	AAAACAACAA	AACAAAAAAA	1330
	CAGAAAGAAA	GAAAACTACA	GGCCAATATC	CCTGATGAAT	ACTGATACAA	AAATCCTCAA	CAAAACACTA	1400
25	GCAAAACCAA	TTAAACAACA	CCCTCGAAG	ATCATTCAIT	GTGATCAAGT	GGGATTTATT	CCAGGGATGG	1470
	AAGGATGGTT	CAACATATGC	AAATCAATCA	ATGTGATACA	TCATCCCAAC	AAAAATGAAGT	ACAAAAACTA	1540
	TATGATTATT	TCACTTTATG	CAGAAAAAGC	ATTTGATAAA	ATTCGACACC	CTTCATGATA	AAACCCCTCA	1610
	AAAAACCAAG	TATACAAGAA	ACATACAGGC	CAGGCACAGT	GGCTCACACC	TGCGATCCCA	GCACTCTGGG	1680
	AGGCCAAGGT	GGGATGATTG	CTTGGGCCCA	GGAGTTTGAG	ACTAGCCTGG	GCAACAAAT	GAGACCTGGT	1750
	CTACAAAAAA	CTTTTAAAAA	AAATTAGCCA	GGCATGATGG	CATATGCCTG	TAGTCCAGC	TAGTCTGGAG	1820
30	GCTGAGGTGG	GAGAATCACT	TAAGCCTAGG	AGGTCGAGGC	TGCAGTGAGC	CATGAACATG	TCACGTACT	1890
	CCAGCCTAGA	CAACAGAACCA	AGACCCCACT	GAATAAGAAG	AAGGAGAAAG	AGAAAGGAGA	AGGGAGGGAG	1960
	AAGGGAGGAG	GAGGAGAAGG	AGGAGGTGGA	GGAGAAAGTG	AAGGGGAAGG	GGAAGGGAAA	GAGGAAGAAG	2030
	AAGAAACATA	TTTCAACATA	ATAAAGCCCC	TATATGACAG	ACCGAGGTAG	TATTATGAGG	AAAAACTGAA	2100
	AGCCTTTCCT	CTAAGATCTG	GAAATGACA	AGGGCCCACT	TTCACCACTG	TGATTCAACA	TAGTACTAGA	2170
35	AGTCTTAGCT	AGAGCAATCA	GATAAGAGAA	AGAAATAAAA	GGCATCCAAA	CTGGAAAGGA	AGAAAGTCAA	2240
	TTATCTCTGT	TGCAGATGAT	ATGATCTTAT	ATCTGGAAAA	GACTTAAAGC	ACCACTAAAA	AACATTTAGA	2310
	GCTGAAATTT	GGTACAGCAG	GATACAAAAA	CAATGTACAA	AAATCAGTAG	TATTTCATATA	TTCCAAACAGC	2380
	AAACAATCTG	AAAAAGAAAC	CAAAAAAGCA	GCTACAAATA	AAATTAACA	GCTAGGAATT	AACCAAGAAA	2450
	GTGAAAGATC	TCTACAATGA	AAACTATAAA	ATGTTGATAA	AAGAAATTGA	AGAGGGCACA	AAAAAAGAAA	2520
40	AGATATTCCA	TGTTTCATAGA	TTGGAAGAAAT	AAATACTGTT	AAATGTCCA	TACTACCCAA	AGCAATTTAC	2590
	AAATTCAATG	CAATCCCTAT	TAAAAATACTA	ATGACGTTCT	TCACAGAAAT	AGAAGAAACA	ATTCTAAGAT	2660
	TTGTACAGAA	CCACAAAAGA	CCGAGAATAG	CCAAAGCTAT	CCTGACCAAA	AAGAACAATA	CTGGAAAGCAT	2730
	CACATTACCT	GACTTCAAAAT	TATACTACAA	AGCTATAGTA	ACCCAAACTA	CATGGTACTG	GCATAAAAAAC	2800
	AGATGAGACA	TGGACCAAG	GAACAGAAATA	GAGAATCCAG	AAACAAATCC	ATGCATCTAC	AGTGAACCTCA	2870
45	TTTTTGACAA	AGGTGCCAAG	AACTACTTTT	GGGAAAAAGA	TAATCTCTTC	AATAAATGGT	CTGGAGGAAA	2940
	CTGGATATCC	ATATGCAAAA	TAAACATACT	AGAACTCTGT	CTCTCACCAT	ATACAAAAGC	AAATCAAAAT	3010
	GGATGAAAGG	CTTAAATCTA	AAACCTCAAA	CTTTGCAACT	ACTAAAAGAA	AACACCGGAG	AAACTCTCCA	3080
	GGACATTGGA	GTGGGCAAGG	ACTTCTTGAG	TAATTCCTTG	CAGGCACAGG	CAACCAAGC	AAAAACAGAC	3150
	AAATGGGATC	ATATCAAGTT	AAAAAGCTTC	TGCCAGCAAA	AGGAACAAT	CAACAAAGAG	AAGACAGAAC	3220
50	CCACAGAAATG	GGAGAAATATA	TTTGCAAACT	ATTCACTTAA	CAAGGAATTA	ATAACCAAGTA	TATATAAGGA	3290
	GCTCAAACTA	CTCTATAAGA	AAAAACCTTA	ATAAGCTGAT	TTTCAAAAAT	AAGCAAAAAG	TCTGGGTAGA	3360
	CATTCTCTAA	AATAAGTCAT	ACAAATGGCA	AACAGGCATC	TGAAATGTG	CTCAACACCA	CTGATCATCA	3430
	GAGAAATGCA	AATCAAAACT	ACTATGAGAG	ATCATCTCAT	CCCAGTTAAA	ATGGCTTTTA	TTCAAAAGAC	3500
	AGGCAATAAC	AAATGCCAGT	GAGGATGTGG	ATAAAGGAA	ACCCTTGGAC	ACTGTTGGTG	GGAATGGAAA	3570
55	TTGCTACCAC	TATGGAGAAC	AGTTTGAAAG	TTCTCAAAA	AACTAAAAAT	AAAGCTACCA	TACAGCAATC	3640
	CCATTGCTAG	GTATATACTC	CAAAAAAGGG	AATCAGTGTA	TCAACAAAGCT	ATCTCCACTC	CCACATTTAC	3710
	TGCAGCACTG	TTATAGCAG	CCAAGGTTTG	GAAGCAACCT	CAGTGTCCAT	CAACAGACGA	ATGGAAAAAG	3780
	AAAAATGGGT	GCACATACAC	AATGGAGTAC	TACGCAGCCA	TAAAAAAGAA	TGAGATCTCTG	TCAGTTGCCAA	3850
	CAGCATGGGG	GGCACTGGTC	AGTATGTTAA	GTGAAATAAG	CCAGGCACAG	AAAGACAAAC	TTTTCATGTT	3920
60	CTCCCTTACT	TGTGGGAGCA	AAAAATTAATA	CAATTGACAT	AGAAATAGAG	GAGAAATGGTG	GTCTAGAGG	3990
	GGTGGGGGAC	AGGGTGACTA	GAGTCAACAA	TAATTTATTG	TATGTTTTAA	AATAACTAAA	AGAGTATAAT	4060
	TGGGTTGTTT	GTAACACAAA	GAAAGGATAA	ATGCTTGAAG	GTGACAGATA	CCCCATTAC	CCTGATGTGA	4130
	TTATTACACA	TTGTATGCCCT	GTATCAAAAT	ATCTCATGTA	TGCTATAGAT	ATAAACCCCTA	CTATATTAAT	4200
	AATTAAAAAT	TTAATGGCCA	GGCAGCGTGG	CTCATGTCCG	TAATCCACAG	ACTTTGGGAG	GGCAGGGCGG	4270
65	GTGGATCACC	TGAGGTCAGG	AGTTTGAATC	CAGTCTGGCC	ACCATGATGA	AACCTGTCT	CTACTAAAAA	4340
	TACAAAAAAT	AGCCAGGCGT	GGTGGCACAT	ACCTGTAGTC	CCAACCTACT	AGGAGGCTGA	GACAGGAGAA	4410
	TTGCTTGAACT	CTGGAGGGCG	GAGGTGTCAG	TGAGCCGAGA	TCATGCCACT	GCACTGCAGC	CTGGGTGACA	4480
	GAGCAAGACT	CCATCTCAAA	ACAAAAACAA	AAAAAAGAAG	ATTAAAAATG	TAATTTTAT	GTACCGTATA	4550
	AATATATACT	CTACTATATT	AGAAGTTAAA	AATTAAAAAA	ATTAAAAAG	GTAATTAACC	ACTTAATCTA	4620
70	AAATAAGAAC	AATGTATGTG	GGGTTTCTAG	CTTCTGAAGA	AGTAAAAATT	ATGGCCACGA	TGGCAGAAAT	4690
	GTGAGGAGGG	AACAGTGGAA	GTTACTGTTG	TTAGACGCTC	ATACTCTCTG	TAAGTGACTT	AATTTTAACC	4760
	AAAGACAGGC	TGGGAGAAGT	TAAAGAGGCA	TTCTATAAGC	CCTAAAACAA	CTGCTAATAA	TGGTGAAGGG	4830
	TAATCTCTAT	TAATTACAGA	TATCTCTAAA	ATCGAGCTGC	AGAAATGGCA	AGAAATGGCA	CTCTGATACA	4900
	CACCGTCTCT	TCATTACAGG	TGCTTTTTTT	CTTGTGTGCT	TGGAGATTTT	CGATTGTGTG	TTGCTGTTTG	4970
75	GTTAAACTTA	ATCTGTATGA	ATGCTGAAAC	GAAAAATGGT	GGTGATTTC	TCCAGAAAGA	TTAGAGTACC	5040
	TGGCAGGAAG	CAGGTGGCTC	TGTGGACCTG	AGCCACTTCA	ATCTTCAAGG	GTCTCTGGCC	AAGACCAGG	5110

	TGCAAGGCAG	AGGCCTGATG	ACCCGAGGAC	AGGAAAGCTC	GGATGGGAAG	GGGCGATGAG	AAGCCTGCCT	5180
	CGTTGGGTGAG	CAGCGCATGA	AGTGCCCTTA	TTTACGCTTT	GCAAAGATTG	CTCTGGATAC	CATCTGGAAA	5250
	AGGCGGCCAG	CGGGAATGCA	AGGAGTCAGA	AGCCTCCTGC	TCAAACCCAG	GCCAGCAGCT	ATGGCGCCCA	5320
	CCCGGGCGTG	TGCCAGAGGG	AGAGGAGTCA	AGGCACCTCG	AAGTATGGCT	TAAATCTTTT	TTTCACCTGA	5390
5	AGCACTGACC	AAGGTGTATT	CTGAGGGAAG	CTTGAGTTAG	GTGCCTTCTT	TAAACAGAA	AGTCATGGAA	5460
	GCACCCCTCT	CAAGGGAATA	CCAGACGCCC	GCTCTGCGGT	CATTTACCTC	TTTCCTCTCT	CCCTCTCTTG	5530
	CCCTGCGGGT	TTCTGATCGG	GACAGAGTGA	CCCCCGTGA	GCTTCTCGGA	GCCCGTGCTG	AGGACCCCTC	5600
	TGCAAGGGG	TCCACAGACC	CCCGCCCTGG	AGAGAGGAGT	CTGAGCCTGG	CTTAATAACA	AACCTGGAGT	5670
	TGGCTGGGGG	CGGACAGCGA	CGCGGGGATT	CAAAAGACTTA	ATTCCATGAG	TAAATTC AAC	CTTTCCACAT	5740
10	CCGAATGGAT	TTGGATTTTA	TCTTAATATT	TTCTTAAATT	TCATCAAATA	ACATTCAGGA	CTGCAGAAAT	5810
	CCAAAGCGGT	AAAACAGGAA	CTGAGCTATG	TTTGCCAAAG	TCCAAGGACT	TAATAACCAT	GTTTCAGAGG	5880
	ATTTTTCGCC	CTAAGTACTT	TTTATTGGTT	TTTATAAGGT	GGCTTAGGGT	GCAAGGGAAA	GTACACGAGG	5950
	AGAGGCTTGG	GCGGCAGGGC	TATGAGCAGC	GCAGGGCCAC	CGGGGAGAGA	GTCCCCGGCC	TGGGAGGCTG	6020
	ACAGCAGGAC	CACTGACCGT	CCTCCCTGGG	AGCTGCCACA	TTGGGCAACG	GCAAGGGGGC	CACCTCGCTG	6090
15	GTGACTCAGG	ACCCCATACC	GGCTTCTGGG	GCCCACCCAC	ACTAACCAG	GAAGTCACGG	AGCTCTGAAC	6160
	CCGTGGAAAC	GAACATGACC	CTTGCTGGCC	TGCTTCCCTG	GGTGGGTCAA	GGGTAATGAA	GTGGGTGCA	6230
	GGAAATGGCC	ATGTAATTTA	CACGACTCTG	CTGATGGGGA	CCGTTCCTTC	CATCATTATT	CATCTTCACC	6300
	CCCAAGGACT	GAATGATTCA	AGCAACTTCT	TCGGGTGTGA	CAAGCCATGA	CAAAACTCTAG	TACAAACACC	6370
	ACTCTTTTAC	TAGGCCACCA	GAGCAGGSC	CACACCCCTG	ATATATTAA	AGTCCAGGAG	AGATGAGGCT	6440
20	GCTTTTCAGC	ACCAGGCTGG	GGTGACAACA	GCGGCTGAAC	AGTCTGTTC	TCTAGACTAG	TAGACCTGG	6510
	CAGGCACCTC	CCCAGATTCT	AGGGCTGGT	TGCTGCTTCC	CGAGGGCCGC	ATCTGCCCTG	GAGACTCAGC	6580
	CTGGGTGGCC	ACACTGAGGC	CAGCCCTGTC	TCCACACCTT	CCGCCCTCAG	GCCTCAGCTT	CTCCAGCAGC	6650
	TTCTTAAACC	CTGGGTGGCC	CGTGTCCAG	CGCTACTGTC	TCACCTGTCC	CACCTGTGCT	TGCTCAGGCT	6720
	ACGTAGCTCG	CACGCTTCTT	CCTCAGATGG	GGTGTCTGTC	TCCTTCCCCA	ACACTCAGAT	CGCTGAAGG	6790
25	GAGGAGATTC	TGCGCCTCCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGGCCCCC	GATGCAGGTT	6860
	CCTGGCTGCC	GGCTGCACGC	TGACCTCCAT	TCCAGGCGC	TCCCCTGCTC	CTGTATCTG	CCGGGGCTG	6930
	CCGGTGTGTT	CTTCTGTTT	TGTGCTCCT	TCCACGTCCA	GCTGCTGTG	TCTCTGCCCT	CTAGGGTCTC	7000
	GGGGTTTTTA	TAGGCATAGG	ACGGGGGCGT	GGTGGGCCAG	GGCGCTCTTG	GGAAATGCAA	CATTTGGGTT	7070
30	TGAAAGTAGG	AGTGCTGTC	CTCACCTAGG	TCCACGGGCA	CAGGCCTGGG	GATGGAGCCC	CCGCCAGGGA	7140
	CCCGCCCTTC	TCTGCCAGC	ACTTCTCTGC	CCCCCTCCCT	CTGGAACACA	GAGTGGCAGT	TCCACAAGC	7210
	ACTAAGCATC	CTCTTCCCAA	AAGACCCAGC	ATTGGCACCC	CTGGACATTT	GCCCCACAGC	CCTGGGAATT	7280
	CAGGTGACTA	CGCACATCAT	GTACACACTC	CCGTCCACGA	CCGACCCCGC	CTGTTTTATT	TAAATAGCTA	7350
	CAAGCAGGGG	AAATCCCTGC	TAAATGTCC	TTTAAACAA	TGTTTAAACA	AACGGGTCCA	TCCGCACGTT	7420
35	GGACAGTTCC	TCACAGTGAA	GAGGAACATG	CCGTTTATAA	AGCCTGCAGG	CATCTCAAGG	GAATTACGCT	7490
	GAGTCAAAC	TGCCACCTCC	ATGGGATACG	TACGCAACAT	GCTCAAAAG	AAGAATTTT	ACCCATGGC	7560
	AGGGGAGTGG	TAGGGGGGTT	TAAGGACGTT	GGGGGCGGCA	GCTGGGGGCT	ACTGCACGCA	CCTTTTACTA	7630
	AAGCCAGTTT	CCTGTTCTG	ATGGTATTGG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	TGGGGATGGG	7700
	GGAAACCCGA	GGCTGTGCCA	TCTTTGCCAT	GCCCGAGTGT	CCTGGGCAGG	ATAATGCTCT	AGAGATGCC	7770
40	ACGTCTTGAT	TCCCCCAAC	CTGTGGACAG	AACCCGCCCC	GCCCCAGGGC	CTTTCAGGTT	GTGATCTCCG	7840
	TGAGGACCTT	GAGGCTGGG	ATCCTTCGGG	ACTACCTGCA	GGCCCGAAAA	GTAATCCAGG	GGTTCGGGA	7910
	AGAGGCGGGC	AGGAGGGTCA	GAGGGGGGCA	GCCTCAGGAC	GATGGAGGCA	GTCAGTCTGA	GGCTGAAAG	7980
	GGAGGGAGGG	CCTCGAGCCC	AGGCCTGCAC	GCGCCTCCAG	AAGCTGGAAA	AAGCGGGGAA	GGGACCTCC	8050
	ACGGAGCCGT	CAGCAGGAAG	GCACGGCTGG	CCCTTAGCCC	ACCAAGGGCC	ATCGTGGACC	TCCGGCTCC	8120
45	GTGCCATAGG	AGGGCACTCG	CGCTGCCCTT	CTAGCATGAA	GTGTGTGGGG	ATTTCAGAG	GAACACGAA	8190
	ACCCATGCAC	TGTGAATCTA	GGATTATTTT	AAAACAAAGG	TTTACAGAAA	CATCCAAGGA	CAGGGCTGAA	8260
	GTGCCTCCGG	GCAAGGGCAG	GCCAGGCACG	AGTGATTTTA	TTTAGCTATT	TTATTTTATT	TACTTACTTT	8330
	CTGAGACAGA	GTTATGCTCT	TGTTGCCCCG	GCTGGAGTGC	AGCGGCATGA	TCTTGGCTCA	CTGCAACCTC	8400
	CGCTCTCTGG	GTTCAAGCAA	TTCTCGTGCC	TCAGCCTCCC	AAGTAGCTGG	GATTTTCAGG	GTGCACCACT	8470
50	ACACCCGGCT	AATTTTGTAT	TTTTAGTAGA	GATGGGCTTT	CACCATGTTG	GTCAAGCTGA	TCTCAAAATC	8540
	CTGACCTCAG	GTGATCCGCC	CACCTCAGCC	TCCCAAAGTG	CTGGGATTAC	AGGCATGAGC	CACCTGACCT	8610
	GGCCTATTTA	ACCATTTTAA	AACCTTCCCT	GGCTCAAGTC	ACACCCACTG	GTAAGGAGTT	CATGGAGTTC	8680
	AATTTCCCTT	TTACTCAGGA	GTTACCCCTC	TTTGATATTT	TCTGTAATTC	TTGCTAGACT	GGGGATACAC	8750
	CGTCTCTTGA	CATATTCACA	GTTTCTGTGA	CCACCTGTTA	TCCCATGGGA	CCCAGTGCAG	GGGCAGCTGG	8820
55	GAGGCTGCAG	GCTTCAGGTC	CCAGTGGGGT	TGCCATCTGC	CAGTAGAAAC	CTGATGTAGA	ATCAGGGGCG	8890
	AAGTGTGGAC	ACTGTCTTGA	ATCTCAATGT	CTCAGTGTGT	GCTGAAACAT	GTAGAAATTA	AAGTCCATCC	8960
	CTCCTACTCT	ACTGGGATTG	AGCCCTTCCC	CTATCCCCCC	CCAGGGGCGC	AGGAGTTCTC	CTCACTGCTG	9030
	TGGAGGAAGG	AATGATACCT	TGTTATTTT	CACCTGCTGG	ACTGAATCCA	CTGTTTCATT	TGTTGGTTTG	9100
60	TTTGTTTGT	TTTGAGAGGC	GGTTTCACTC	TTGTGTCTCA	GGCTGGAGGG	AGTGCAATGG	CGCGATCTTG	9170
	GCTTACTGCA	GCCTCTGCCT	CCAGGTTTCA	AGTGATTCTC	CTGCTTCCGC	CTCCCATTTG	GCTGGGATTA	9240
	CAGGCACCCG	CCACCATGCC	CAGCTAATTT	TTTGATTTTT	TAGTAGAGAC	GSGGGTGGGT	GGGGTTCACC	9310
	ATGTTGGCCA	GGCTGGTCTC	GAACCTTCTG	CCTCAGATGA	TCCACCTGCC	TCTGCCTCCT	AAAGTGCTGG	9380
	GATTACAGGT	GTGAGCCACC	ATGCCACGCT	CAGAATTTAC	TCTGTTTAGA	AACATCTGGG	TCTGAGGTAG	9450
	GAAGCTCACC	CCACTCAAGT	GTTGTGGTGT	TTTAAAGCAA	TGATAGAATT	TTTTTATTGT	TGTTAGAAC	9520
65	CTCTTGATGT	TTTACACTGT	GATGACTAAG	ACATCATCAG	CTTTTCAAAG	ACACACTAAC	TGCACCCATA	9590
	ATACTGGGGT	GTCTTCTGGG	TATCAGCAAT	CTTCATTGAA	TGCCGGGAGG	CGTTTCTCTG	CCATGCACAT	9660
	GGTGTTAATT	ACTCCAGCAT	AATCTTCTGC	TTCCATTCTC	TCTCTTCCCT	CTTTTAAAT	TGTTTCTTCT	9730
	ATGTTGGCTT	CTCTGCAGAG	AACCAAGTGA	AGCTACAAC	TAACTTTTGT	TGGAACAAAT	TTTCCAAACC	9800
	GCCCTTTTGC	CCTAGTGGCA	GAGACAATTC	ACAAACACAG	CCCTTTAAAA	AGGCTTAGGG	ATCACTAAGG	9870
70	GGATTCTTAG	AAGAGCGACC	TGTAATCCTA	AGTATTTACA	AGACGAGGCT	AACCTCCAGC	GAGCGTGACA	9940
	GCCAGGGGAG	GGTGCGAGGC	CTGTTCAAAT	GCTAGCTCCA	TAAATAAAGC	AATTTCTCTC	GGCAGTTTCT	10010
	GAAAGTAGGA	AAGGTTACAT	TTAAGGTTGC	GTTTGTAGC	ATTTAGAGT	TTGCCAGCT	CAGCTACAGC	10080
	ATCCCTGCAG	GGCCTCGGGA	GACCCAGAAG	TTTCTCGCCC	CCTTAGATCC	AAACTTGAGC	AGCCCGGAGT	10150
	TGCGATTCTT	GGGAAGTCTT	CAGCTGTCTT	GCGGTTGTGC	CGGGGCCCCA	GGTCTGGAGG	GACCCAGTGG	10220
	CCGTGTGGCT	TCTACTGCTG	GGCTGGAAGT	CGGGCTCTCT	AGCTCTGACG	TCCGAGGCTT	GAGGCCAGCT	10290
75	GCCTGGACCC	CGAGGCTGCC	CTCCACCCTG	TGCGGGGCGG	ATGTGACACG	ATGTTGGGCT	CATCTGCGAG	10360
	ACAGAGTGCC	GGGGCCACGG	GTCAGGCCG	TGTGGCTG	TGTGAGGCGC	CCGGTGGCGC	GCCAGCAGGA	10430
	GCGCCTGGCT	CCATTTCCTA	CCCTTCTCTG	ACGGGACCGC	CCCGTGGGTT	GATTAAACAG	TTTGGGGTGG	10500

	TTTGCTCATG	GTGGGGACCC	CTCGCCGCC	GAGAACCTGC	AAAGAGAAAT	GACGGGCTG	TGTCAAGGAG	10570
	CCCAAGTCCG	GGGGAAGTGT	TGCAGGAGG	CACTCCGGGA	GGTCCCGCGT	GCCTGCTCCAG	GGAGCAATGC	10640
	GTCTCGGGT	TCGTCCCCAG	CGCGGTCTAC	GGCCCTCCGT	CCTCCCCCTC	ACGTCCGGCA	TTCTGGTGC	10710
	CCGGAGCCCG	ACGCCCCCGG	TCCGGACCTG	GAGGAGCCCG	TGGGTCTCCG	GATCAGGCCA	CGCGCCAAAG	10780
5	GGTCGCCGCA	CGACCTGTT	CCCAGGGCCT	CCACATCATG	GCCCTCCCTC	CGGGTTACCC	CACAGCCTAG	10850
	GCCGATTTCGA	CCTCTCTCCG	CTGGGGCCCT	CGCTGGCGTC	CCTGCACCTT	GGGAGCCGCA	CGCGCCGCGC	10920
	GGCGGGGAAG	CGCGGCCAG	ACCCCCGGGT	CCGCCCGGAG	CAGCTGCGCT	GTCCGGGCCA	GGCCGGGCTC	10990
	CCAGTGGATT	CGCGGGCACA	GACGCCCAGG	ACCGCGCTCC	CCACGTGGCG	GAGGGACTGG	GGACCCGGGT	11060
	ACCCGTCTCTG	CCCCCTTACC	TTCCAGCTCC	GCCTCTCTCC	CGCGGAGCCC	GCCTCGTCCC	GACCCCTCCC	11130
10	GGGTCCCGCG	CCCAGCCCCC	TCCGGGCCCT	CCCAGCCCTT	CCCCCTTCTT	TCCCGGGCCC	CGCCCTCTCC	11200
	TGCGGGCCCG	AGTTTCAGGC	AGCGCTGCGT	CCTGCTGCGC	ACGTGGGAAG	CCCTGGCCCC	GGCCACCCCC	11270
	GCGATGCCCG	GCGCTCCCGG	CTGCCGAGCC	GTGCGCTCCC	TGCTGCCGAG	CCACTACCCG	GAGGTGCTGC	11340
	CGCTGGCCAC	GTTCGTGCGG	CGCCTGGGGC	CCCAGGGCTG	CGCGCTGGTG	CAGCGCGGGG	ACCCGGCGGC	11410
	TTTCGCCCGG	CTGCTGGCCC	AGTGCCCTGGT	GTGCGTGCCC	TGGGACGCAC	GGCCGCCCCC	CGCCGCCCCC	11480
15	TCCTTCCGCC	AGGTGGGCTT	CCCCGGGGTC	GGCGTCCGGC	TGGGGTTGAG	GGCGGCCGGG	GGGAACACG	11550
	GACATGCGGA	GAGCAGCGCA	GGCGACTCAG	GGCGCTTCCC	CCGCAGGTGT	CCTGCCTGAA	GGAGCTGGTG	11620
	GCCCGAGTGC	TGCAGAGGCT	GTGCGAGCGC	GGCGGAAGA	ACGTGCTGGC	CTTCGGGTTT	GGCTGTGTTG	11690
	ACCGGGCCCG	CGGGGGGCCCT	CCCGAGGCCCT	TCACCAACAG	CGTGCGCAGC	TACCTGCCCCA	ACACGGTGAC	11760
	CGACGCACTG	CGGGGGAGCG	GGGCGTGCGG	GCTGCTGCTG	CGCCCGGTGG	GCGACGACGT	GCTGTTTAC	11830
20	CTGCTGGCAC	GCTGCGCGCT	CTTTGTGCTG	GTGGCTCCCA	GCTGCGCCTA	CCAGGTGTGC	GGGCGCGCGC	11900
	TGTACCACTG	CGCGCTGCCC	ACTCAGGCC	GGCCCCCGCC	ACACGCTAGT	GGACCCCGAA	GGCGTCTGGG	11970
	ATGCGAACGG	GCCTGGAACC	ATAGCGTCAG	GGAGGCCGGG	GTCCCCCTGG	GCCTGCCAGC	CCCCGGTGCG	12040
	AGGAGCGCGG	GGGCGAGTGC	CAGCCGAAGT	CTGCGCTTGC	CCAAGAGGCC	CAGGCGTGGC	GCTGCCCTTG	12110
	AGCCGGAGCG	GACGCCCGTT	GGGCAAGGGT	CCTGGGCCCA	CCCGGGCAGG	ACGCGTGGAC	CGAGTGACCG	12180
25	TGCTTTCTGT	GTGGTGTAC	CTGCCAGACC	CGCCGAAGAA	GCCACCTCTT	TGGAGGGTGC	GCTCTCTGGC	12250
	ACCGGCCACT	CCCCACCATC	CGTGGGCCCG	CAGCACACAG	CAGGCCCCCC	ATCCACATCG	CGGCCACACG	12320
	GTCCCTGGGA	CACGCTTGT	CCCCCGGTGT	ACGCCGAGAC	CAAGCACTTC	CTCTACTCTT	CAGGCGACAA	12390
	GGAGCAGCTG	CGGCCCTCCT	TCCTACTCAG	CTCTCTGAGG	CCACGCTGA	CTGGCGCTCG	GAGGCTCGTG	12460
	GAGACCATCT	TTCTGGGTTG	CAGGCCCTTG	ATGCCAGGGA	CTCCCGCAG	GTTCGCCCGC	CTGCCCCAGC	12530
30	GCTACTGGCA	AATGCGGCC	CTGTTTCTGG	AGCTGCTTGG	GAACCAACGG	CAGTGCCCTT	ACGGGGTGTCT	12600
	CCTCAAGACG	CACTGCCCGC	TGCGAGCTGC	GGTCAACCCA	GCAGCCGGTG	TCTGTGCCCG	GGAGAAGCCC	12670
	CAGGGCTCTG	TGGCGGGCCC	CGAGGAGGAG	GACACAGACC	CCCGTCGCGT	GCTGCACTGT	CTCCGCCAGC	12740
	ACAGCAGCCC	CTGGCAGGTG	TACGGCTTCG	TGCGGGCCTG	CCTGCGCCGG	CTGTTGCCCG	CGGGCTCTGT	12810
35	GGGCTCCAGG	CACAACGAAC	CGCGCTTCTT	CAGGAACACC	AAGAAGTTCA	TCTCCCTGGG	GAAGCATGCC	12880
	AAGCTCTCCG	TGCAGGAGCT	GACGTGGAAG	ATGAGCGTGC	GGGACTGCGC	TGGCTGCGC	AGGAGCCAG	12950
	GTGAGGAGGT	GGTGGCCGTC	GAGGGCCAG	GCCCAGAGC	TGAATGCAGT	AGGGGCTCAG	AAAGGGGGCG	13020
	AGGCGAGGCC	CTGGTCTCTC	TGCTTCCATC	CTACAGTGGG	CACACGTGGC	TTTTCGCTCA	GGAGCTCAG	13090
	TGGACACGGT	GATCTCTGCC	TCTGCTCTCC	CTCCTGTCCA	GTTTGCAATA	ACTTACGAGG	TTACCTTCA	13160
40	CGTTTTGATG	GACACGCGGT	TTCCAGGCGC	CGAGGCCAGA	GCAGTGAACA	GAGGAGCTGT	GGCGCGGCG	13230
	TGGAGCCGGG	TTGCCGGCAA	TGGGGAGAAG	TGCTTGAAG	CACAGACGCT	CTGGCGAGGG	TGCCCTGACG	13300
	TTACCTATAA	TCTCTTCCG	AATTTCAAGG	GTGGGAATGA	GAGGTGGGGA	CGAGAACCCC	CTCTTCTTGG	13370
	GGGTGGGAGG	TAAGGGTTTT	GCAGGTGCAC	GTGGTCAGCC	AATATGCAGG	TTTGTGTTTA	AGATTTAATT	13440
	GTGTGTTGAC	GGCCAGGTGC	GGTGGCTCAC	GGCGGTAAATC	CCAGCACTTT	GGGAAGCTGA	GGCAGGTGGA	13510
45	TCACCTGAGG	TCAGGAGTTT	GAGACACGCC	TGACCAACAT	GGTGAACACC	TATCTGTACT	AAAAATACAA	13580
	AAATTAGCTG	GGCATGGTGG	TGTGTGCCCTG	TAATCCACAG	TACTTGGGAG	GCTGAGGCAG	GAGAATCACT	13650
	TGAACCCAGG	AGGCGGAGGC	TGCAGTGAGC	TGAGATTGTG	CCATTGTACT	CCAGCCTGGG	CGACAAGAGT	13720
	GAAACTCTGT	CTTTAAAAAA	AAAAAGTGT	CGTTGATTGT	GCCAGGACAG	GCTAGAGGGA	GGGAGATAAG	13790
	ACTGTTCTCC	AGCAGAGATC	CTGGTCCCAT	CTTTAGGTAT	GAAGAGGGCC	ACATGGGAGC	AGAGGACAGC	13860
50	AGATGGCTCC	ACCTGCTGAG	GAAGGGACAG	TGTTTGTGGG	TGTTCAAGGG	ATGGTGCTGC	TGGGCCCTGC	13930
	CGTGTCCCA	CCCTGTTTTT	CTGGATTGTA	TGTTGAGGAA	CCTCCGCTCC	AGCCCCCTTT	TGGCTCCAGC	14000
	TGCTCCACAG	CCCTACCGTG	GCAGCTAGAA	GAAGTCCCGA	TTTCAACCCC	TCCCAACAAA	CTCCCAAGAC	14070
	ATGTAAGACT	TCCGGCCATG	CAGACAAGGA	GGGTGACCTT	CTTGGGGCTC	TTTTTTTTCT	TTTTTTCTTT	14140
	TTATGGTGGC	AAAAGTCATA	TAACATGAGA	TTGGCACTCC	TAACACCGTT	TTCTGTGTAC	AGTGCAGAAT	14210
55	TGCTAACTCG	GCGGTGTTTA	CAGCAGGTTG	CTTGAATATG	TGCGTCTTGC	GTGACTGGAA	GTCCCTACCC	14280
	ATCGAACGGC	AGCTGCCCTCA	CACCTGCTGC	GGCTCAGGTG	GACCACGCCG	AGTCAGATAA	CGCTCATGCA	14350
	ACCCAGTTTT	GCTTTTTTGTG	CTCCAGCTTC	CTTCGTTGAG	GAGAGTTTGA	GTTCTCTGAT	CAGGACTCTG	14420
	CCTGTCAATTG	CTGTTCTCTG	ACTTCAGATG	AGGTCACAAT	CTGCCCTTGG	CTTATGCAGG	GAGTGAGCGG	14490
	TGGTCCCCCG	GTGTCCTGT	CACGTGACAG	GTGAGTGAGG	CGTTGCCCCC	AGGTGTCCCT	GTCACGTGTA	14560
60	GGGTGAGTGA	GGCGCGGCC	CCGGGTGTCC	CTGTCCCGTG	CAGCGTGATT	GAGGTGTGGC	CCCCGGGTGT	14630
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCGCC	ATCCCCGGGT	GTCCCTGTCA	CGTGTAGGGT	GAGTGAGCGG	14700
	TGGTCCCCCG	GTGTCCCTGT	CCCGTGCAGG	GTGAGTGAGG	CACTGTCCCC	GGGTGTCCCT	GTCACGTGCA	14770
	GGGTGAGTGA	GGCGCGGTC	CCGGGTGTCC	CTCTCAGGTG	TAGGGTGAGT	GAGGCGCGGC	CCCAGGGTGT	14840
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCGCC	GTCCCTGGGT	GTCCCTCCCA	GGTATAGGGT	GAGTGAGGCA	14910
65	CTGTCCCCCG	GTGTCCCTGT	CACGTGACAG	GTGAGTGAGG	CGCGGCCCCC	GGGTGTCCCT	CTCAGGTGCA	14980
	GGGTGAGTGA	GGCGCTGTCC	CTGGGTGTCC	CTGTCTCTGT	TAGGGTGAGT	GAGGCTCTGT	CCCCAGGTGT	15050
	CCTTGGCGTT	TGCTCACTTG	AGCTTGCTCC	TGAATGTTTG	CTCTTTCTAT	AGCCACAGCT	GGCGCGGTTG	15120
	CCCATTGCTT	GGGTAGATGG	TGCAGGCGCA	GTGCTGGTCC	CCAAGCCTAT	CTTTCTGAT	GCTCGGCTCT	15190
	TCTTGGTCA	CTCTCCGTTT	CATTTTGCTA	CGGGGACACG	GGACTGCAGG	CTCTCGCCTC	CCGCGTGCCA	15260
70	GGCACTGCAG	CCACAGCTTC	AGGTCCGCTT	GCCTCTGTTG	GGCCTGGCTT	GCTCACCACG	TGCCCGCCAC	15330
	ATGCATGCTG	CCAATACTCC	TCTCCACAGT	TGTCTCATGC	CGAGGCTGGA	CTCTGGGCTG	CTGTGTCGTG	15400
	CTGCCACGTG	TTGCTGGAGA	CATCCAGAAA	AGGGTCTCT	GTGCCCTGAA	GGAAAGCAAG	TCACCCACAG	15470
	CCCCCTCACTT	GTCTGTTTTT	CTCCCAAGCT	GCCCTCTGCT	TGGGCCCCCT	TGGGTGGGTG	GCACCGCTTG	15540
	TCACCTTATT	CTGGGACCTT	GGCGCTCATT	GCTTAGGCTG	GGCTCTGCTT	CCAGTCGCCG	CCTCACATGG	15610
75	ATTGACGTCC	AGCCACAGGT	TGGAGTGTCT	CTGTCTGTCT	CCTGCTCTGA	GACCCACGTG	GAGGGCCGGT	15680
	GTCTCCGCCA	GCCTTCGTCA	GACTTCCCTC	TGGGCTCTTA	GTTTTGAATT	TCAGTGATT	ACCTCTGACG	15750
	TTTCTATCTC	TCCATTGTAT	GCTTTTTCTT	GGTTTATTCT	TTTATTCTCT	TTCTAGCTTC	TTAGTTTAGT	15820
	CATGCCCTTC	CCTCTAAGTG	CTGCCCTTACC	TGCACCCTGT	GTTTTGATGT	GAAGTAATCT	CAACATCAGC	15890

	CACTTTCAAG	TGTTCTTAAA	ATACITCAAA	GTGTTAATAC	TTCTTTTAAAG	TATTCTTATT	CTGTGATTTT	15960
	TTTCTTTGTG	CACGCTGTGT	TTTGACGTGA	AATCATTITG	ATATCAGTGA	CTTTTAAAGTA	TTCTTTAGCT	16030
	TATTCGTGTA	TTTCTTTGAG	CAGTGAGTTA	TTGAACACT	GTITATGTTC	AAGATATGTA	GAGTATCAAG	16100
	ATACGTAGAG	TATTTTAAAGT	TATCATTITA	TTATTGATTT	CTAACTCAGT	TGTGTAGTGG	TCTGTATAAT	16170
5	ACCAATTATT	TGAAGTTTGC	GGAGCCTTGC	TTTGATGATC	AGTGTGTGCA	TGGTTTCCAG	AACTGTCCAT	16240
	TGTTAAATTTG	ACATCCTGTG	AATAGTGGGC	ATGCATGTTC	ACTATATCCA	GCTTATTAAG	GTCCAGTCCA	16310
	AAGCTTCTGT	CTCCTTCTAG	ATGCATGAAA	TTCCAAGAAAG	GAGGCCATAG	TCCCTCACCT	GGGGGATGGG	16380
	TCTGTTTCAAT	TCTTCTCGTT	TGGTAGCATT	TATGTGAGGC	ATTGTTAGGT	GCATGCACGT	GGTAGAATTT	16450
10	TTATCTTCCT	GATGAGTGAA	CTTTTGGAG	ACTTCTATGT	CTCTAGTAAT	CTAGTAATTC	TTTTTTTAAA	16520
	TTGCTCTTAG	TACTGCCACA	CTGGGCTTCT	TTTGATTAGT	ATTTTCTCTG	TGTGTCTGTT	TTCTGCCTTT	16590
	AATTTATATA	TATATATATA	TTTTTTTTTT	TTTTGAGACA	GAGTCTTGGT	CTGTCCGCCA	GGGTGAGTGC	16660
	AGTGGTGTGA	TCACAGGTCA	GTGTAACITT	TACCTTCTGG	CCTGAGCCGT	CCTCTCACCT	CAGCCTCCTG	16730
	AGTAGCTGGA	ACTGCAGACA	CGCACCGCTA	CACCTGGCTA	ATTTTAAAT	TTTTTCTGGA	GACAGGGTCT	16800
15	TGCTGTGTG	CCCAGGCTGG	TCTCAAACCT	TTGACTCAA	GGGATCCATC	TACCTCGGCT	TCCCAAAGTG	16870
	CTGAATTACA	GGCATGAGCC	ACCATGTCTG	GCCTAATTTT	CAACACTTTT	ATATTCTTAT	AGTGTGGGTA	16940
	TGTCCTGTGA	ACAGCATGTA	GGTGAATTTT	CAATCCAGTC	TGACAGTCGT	TGTTTAACTG	GATAACCTGA	17010
	TTTATTTTCA	TTTTTTTGTG	ACTAGAGACC	CGCCTGGTGC	ACTCTGATTC	TCCACTTGCC	TGTTGCATGT	17080
	CCTCGTTCCC	TTGTTTCTCA	CGACCTCTTG	GGTTGCCATG	TGCGTTTCTT	GCCGAGTGTG	TGTTGATCCT	17150
20	CTCGTTGCCT	CCTGGTCACT	GGGCATTGCG	TTTTATTCTT	CTTTGCTTAG	TGTTACCCCC	TGATCTTTTT	17220
	ATTGTCGTTG	TTTGCTTTTG	TTTATTGAGA	CAGTCTCACT	CTGTCACCCA	GGCTGGAGTG	TAATGGCACA	17290
	ATCTCGGCTC	ACTGCAACCT	CTGCCCTCCT	GGTTCAAGCA	GTTCCTATTG	CTCAACCTCA	TGAGTAGCTG	17360
	GGATTACAGG	CGCCCAACAC	CACGCTGGGC	TAATTTTTGT	ATTTTATAGTA	GAGATAGGCT	TTCCACATGT	17430
	TGGCAGGCT	GGTCTCAAAC	TCCAGACCTC	AAGTGATCTG	CCCGCCTTGG	CCTCCACAG	TGCTGGGATT	17500
25	ACAGGTGCAA	GCCACCGTGC	CCGGCATACC	TTGATCTTTT	AAAAATGAAGT	CTGAAACATT	GCTACCCCTG	17570
	TGCTGAGCAA	TAAGACCTTT	AGTGTATTTT	AGCTCTGGCC	ACCCCCAGC	CTGTGTGCTG	TTTTCCCTGG	17640
	TGACTTAGTT	CTATCTCAGG	CATCTTGACA	CCCCCACAAG	CTAAGCATT	TAAATATTGT	TTTCCGTGTT	17710
	GAGTGTCTT	GTAGCTTTGC	CCCCGCCCTG	CTTTTCTCC	TTTTTTCCCT	GTCTGTCTTC	TGCTCTCAGG	17780
30	CCGCGTCTG	GGGTCCCTCT	CCTTGTCTCT	TGCGTGGTTC	TTCTGTCTTG	TTATTGCTGG	TAAACCCAG	17850
	CTTTACCTGT	GCTGGCTTCC	ATGGCATCTA	GGCAGCTCCG	GGGACCTCTG	CTTATGATGC	ACAGATGAG	17920
	ATGTGGAGAC	TCACGAGGAG	GGCGGTCACT	TTGGCCCTGG	AGTGTCTGGA	GCACCCAGTG	GCCAGCTTTC	17990
	CTTAGCCAGT	GAGTGACAGC	AACGTCCGCT	CGGCTGGGT	TCAGCCTGGA	AAACCCAGG	CATGTCCGGG	18060
	TCTGGTGGCT	CCGCGGTGTC	GAGTTTGAAG	TCGCGCAAC	CTGCGGTGTC	GCGCCAGCTC	TGACGGTGCT	18130
	GCCTGGCGGG	GGAGTGTCTG	TTTCTTCCCT	TCTGTTGGG	AACAGGACA	AAGGATGAGG	TCCGAGCCG	18200
35	TTGTGCGCCA	ACAGGAGCAT	GACGTGAGCC	ATGTGGATAA	TTTTTAAAT	TCTAGGCTGG	GCGCGGTGGC	18270
	TCACGCCCTG	AATCCAGCA	CTTTGGGAGG	CCAAGCGGGG	TGGATCACGA	GGTCAGGAGG	TGAGACCAT	18340
	CCTGGCCAAC	ATGATGAAAC	CCCATCTGTA	CTAAAAACAC	AAAAATTAG	TGGCGCTGGT	GGCGGGTGCC	18410
	TGTAATCCCA	GCTACTCGGG	AGGCTGAGGC	AGGAGAATTG	CTTGAACCTG	GGAGTTGGAA	GTTCAGTGA	18480
	GCCGACATTG	CACCACTGCA	CTCCAGCCTG	GCAACACAGC	GAGACTCTGT	CTCAAAAAAA	AAAAAAAAAA	18550
40	AAAAAAAATA	AATTTCTAGTA	GCCACATTAA	AAAAGTAAAA	AAGAAAGGT	GAAATTAATG	TAATAATAGA	18620
	TTTTACTGAA	GCCGAGCAT	TCCACACCTC	ATCATTTTAG	GGTGTATTG	GTGGGAGCAT	CACACACAG	18690
	CAATTGACAA	TTTTTTGAGC	TTTGTCTGCG	GGATCCCGTG	TGTAGGTCCC	GTGCGTGGCC	ATCTCGGCTC	18760
	GGACCTGCTG	GGCTTCCCAT	GGCCATGGCT	TTGTACCAG	ATGGTGCAGG	TCCGGGATGA	GGTCGCCAGG	18830
	CCCTCAGTGA	GCTGGATGTG	CAGTGTCCGG	ATGGTGCAGC	TCTGGGATGA	GGTCGCCAGG	CCCTGCTGTG	18900
45	AGCTGGATGT	GTGGTGTCTG	GATGGTGCAG	GTGAGGGGTG	AGGTCTCCAG	GCCCTCGGTG	AGCTGGAGGT	18970
	ATGGAGTCCG	GATGATGCAG	GTCCGGGGTG	AGGTCCGCAG	CCCCTGCTGT	GAGCTGGATG	TGTGGTGTCT	19040
	GGATGGTGCA	GGTCAGGGGT	GAGGTCTCCA	GGCCCTCGGT	AGCTGGAGG	TATGGAGTCC	GGATGATGCA	19110
	TGAGGTCAAC	AGGCGCTGCG	GTGAGCTGGG	TGTGCGGTGT	CTGGATGGTG	CAGGTCTGGA	GTGAGGTGCG	19180
50	CAGACGGTGC	CAGACCATGC	GGTGAGCTGG	ATATGCGGTG	TCCGGATGGT	CGAGGTCTGG	GGTGAGGTG	19250
	CCAGGCCCTG	CTGTGAGTTC	GATGTGGGGT	GTCCGGATGC	TGCAGGTCCG	GTGTGAGGTC	ACCAGGCCCT	19320
	GCTGTGAGCT	GGATGTGTGG	TGCTTGGATG	GTGCAAGTCT	GGGGTGAAGG	TCCGCCAGGC	CCTGCTTTTG	19390
	AGCTGGATGT	GTGGTGTCTG	GATGGTGCAG	GTCTGGAGTG	AGGTCCGCAG	GCCCTCGGTG	AGCTGGATGT	19460
	GCAGTGTCCA	GATGGTGCAG	GTCCGGGGTG	AGGTCCGCAG	ACCCTGCGGT	GAGCTGGATG	TGCGGTGTCT	19530
55	GGATGGTGCA	GGTCTGGAGT	GAGGTCCGCA	GGCCCTCGGT	GAGCTGGATG	TATGGAGTCC	GGATGGTGCC	19600
	GGTCCGGGGT	GAGGTCCGCA	GACCTGCTG	TGAGCTGGAT	GTGCGGTGTC	TGGATGGTAC	AGGTCTGGAG	19670
	TGAGGTCCGC	AGACCTTGCT	GTGAGCTGGA	TATGCGGTGT	CCGGATGGTG	CAGGTACAGG	GTGAGGTCTC	19740
	CAGGCCCTCG	GTGAGCTGGA	GGTATGGAGT	CCGGATGATG	CAGGTCCGGG	GTGAGGTCCG	CAGGCCCTGC	19810
	TGTGAACCTG	ATGTCCGGCG	TCTGGATGGT	GCAGGTCTGG	GGTGTGGTCG	CCAGGCCCTC	GGTGAGCTCG	19880
60	AGGTATGGAG	TCCGGATGAT	GCAGGTCCGG	GGTGAGGTTC	CCAGGCCCTG	CTGTGAGCTG	GATGTGCGGC	19950
	GTCTGGATGG	TGCAGGTCTG	GGGTGTGGTC	GCCAGGCCCT	CGGTGAGCTG	GAGGTATGGA	GTCCGGATGA	20020
	TGCAAGTCCG	GGGTGAGGTT	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGCAAGTCC	20090
	GGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGCAAGTCT	GGGTGAGGTC	20160
	CACCAGGCCCT	TGCGGTGAGC	TGTTTGTGCG	GTGTCCGGTT	GCTGAGGTC	CGGGGTGAGT	TGCGCAGGCC	20230
65	CTCGGTGAGC	TGGATGTGCG	GTGTCCCGGT	GTCCGGATGG	TGCAGGTCCA	GGGTGAGGTC	GCTAGGCCCT	20300
	TGGTGGGCTG	GATGTGCCCT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	20370
	GATGTGCCCT	GTCTGCATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	GATGTGGGTC	20440
	GTCCGGATGG	TGCAGGTCCG	CGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCGG	TGTCTGGATG	20510
70	GTCCAGGTGC	GGGTGAGGTC	AGCCAAGGCC	TTCCGTGAGC	TGGATGTGGG	GTGTCCGGAT	GGTGAGGTC	20580
	CGGGGTGAGG	TGCGCAGGCC	CTCGGTTAG	CTGGATATGC	GGGTCCGGA	TGGTGCAGGT	CCGGGTGAGG	20650
	GTCACCAAGC	CCTGCGGTTA	GCTGGATGTG	CGGTGTCTGG	ATGGTGCAGG	TCCGGGGTGA	GGTCCGCAAG	20720
	CCCTGCTGTG	AGCTGGATGT	GCTGTATCCG	GATGGTGCAG	GTCCGGGGTG	AGGTCCGCAAG	CCCTGCAAGT	20790
	GAGCTGGATG	TGCTGTATCC	GGATGGTGCA	GGTCTGGCGT	GAGGTGCGCA	GGCCCTGCGG	TTAGCTGGAT	20860
	ATGCGGTGTC	GGATGGTGCA	GGTCCGGGGT	GAGGTGCGCA	GGCCCTGCGG	TTAGCTGGAT	TGCGGCTGTC	20930
75	CGATGGTGC	AGGTCTGGGG	TGAGGTCCGC	AGGCCCTGCT	GTGAGCTGGA	TGTGCTGTAT	CCGGATGGGT	21000
	CAGGTCCGGG	GTGAGGTGCG	CAGGCCCTGC	GGTGAGCTGG	ATGTCGTGTA	TCCGGATGGT	GCAGGTCTGG	21070
	CGTGAGGTGC	CCAGGCCCTG	CGGTGAGCTG	GATGTGCAGT	GTACGGATGG	TGCAGGTCCG	GGGTGAGGTC	21140
	GCCAGGCCCT	CGCGTGGGCT	GTATGTGTGT	TGCTGGATG	GTGCAAGTTC	GGGGTGAAGT	GCCAGGCCCT	21210

	TGCGGTGAGC	TGGATGTGTG	GTGTCTGGAT	GCTGCAGGTC	CGGGGTGAGT	TCGCCAGGCC	CTCGGTGAGC	21350
	TGGATATGCG	GTGTCCCCGT	GTCCGAATGG	TGCAGGTCCA	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	21420
	GATCTGCCGT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGAAGTC	GATGTGCGGT	21490
5	GTCCGGATGG	TGCAGGTCCG	GGGTGAGGTC	ACCAGGCCCT	CGGTGATCTG	GATGTGGCAT	GTCTTCTCG	21560
	TTTAAGGGGT	TGGCTGTGTT	CGGCCGCGAG	AGCACCCCTC	CGGTGAGGAG	ATCCTGGCCA	AGTTCCTGCA	21630
	CTGGCTGATG	AGTGTGTACG	TCGTGAGAGT	GCTCAGGTCT	TTCTTTTATG	TCACGGAGAC	CACGTTCCTG	21700
	AAGAACAGCG	TCTTTTCTTA	CCGGAAGAGT	GTCTGGAGCA	AGTTGCAAAAG	CATTGGAATC	AGGTACTGTA	21770
	TCCCCACGCC	AGGCCCTCTG	TTCTCGAAGT	CCTGGAACAC	CAGCCCGGCC	TCAGCATGCG	CCTGTCTCAA	21840
10	CTTGCCCTGTG	CTTCCCTGGC	TGTGCAAGTC	TGGGCTGGGA	GCCAGGGGCC	CCGTGACAGG	CCTGGTCCAA	21910
	GTGGATTCTG	TGCAAGGCTC	TGACTGCCTG	GAGCTCACGT	TCTCTTACTT	GTAAATCAG	GAGTTGTGTC	21980
	CAAGTGGTCT	CTAGGGTTTG	TAAAGCAGAA	GGGATTAAAA	TTAGATGGAA	ACACTACCAC	TAGCCCTCCT	22050
	GCCTTTCCCT	GGGATGTGGG	TCTGATTCTC	TCTCTCTTTT	TTTTTCTTTT	TTTGAGATGG	AGTCTCACTC	22120
	TGTTGCCAG	GCTGGAGTGC	AGTGGCATAA	TCTTGCTCA	CTGCAACCTC	CACCTCCTGG	GTTAAGCGA	22190
15	TTACACAGCC	TCAGCCTCCT	AAGTAGCTGG	GATTACAGGC	ACCTGCCACC	ACGCTTGCTG	AAATTTTGTG	22260
	CTTTTAGGAG	AGACGGGGTT	TCACCATGTT	GGCCAGGCTG	GTCTCGAACT	CATGACCTCA	GGTGATCCAC	22330
	CCACCTTGCC	CTCCCAAAGT	GCTGGGTTTA	CAGGCTAAGC	CACCGTGCCC	AGCCCCCGAT	TCTCTTTTAA	22400
	TTCATGCTGT	TCTGTATGAA	TCTTCAATCT	ATTGGAATTA	GGTCATGAGA	GGATAAAATC	CCACCCACTT	22470
	GGCGACTCAC	TGCAGGGAGC	ACCTGTGCAG	GGAGCACCTG	GGGATAGGAG	AGTTCCACCA	TGAGCTAACT	22540
20	TCTAGGTGGC	TGCATTGTAA	TGGCTGTGAG	ATTTTGTCTG	CAATGTTTCGG	CTGATGAGAG	TGTGAGATTG	22610
	TGACAGATTG	AAGCTGGATT	TGCATCAGTG	AGGACGGGGA	CGCGTGGTCT	GGGAGATGCC	AGGCTGGGCT	22680
	AGCCCAAGGC	ATGGTATTAG	CTTCTCCGTG	TCCCGCCAG	GCTGACTGTG	GAGGGCTTTA	GTGAGAGATG	22750
	CAGGGCTTCC	CCAGCTCCCC	TGCACACTCG	AGTCCCTGGG	GGGCTTGTG	ACACCCCATG	CCCAAAATCA	22820
	GGATGTCTGC	AGAGGGAGCT	GGCAGCAGAC	CTCCTCAGAG	GTAACACAGC	CTCTGGGCTG	GGGACCCCGA	22890
25	CGTGGTGCTG	GGGCCATTTC	CTTGCATCTG	GGGAGGGGTC	AGGGCTTTC	CTGTGGGAAC	AAGTAAATAC	22960
	ACAAATGCACC	TTACTTAGAG	TTTACACGTA	TTTAATGGTG	TGCGACCCAA	CATGGTCATT	TGACCAATAT	23030
	TTTGGAAAGA	ATTTAATTGG	GGTGACCGGA	AGGAGCAGAC	AGACGTGGTG	GTCCCCAAGA	TGCTCTCTGT	23100
	CACACTCTGG	ACTGTTGTTC	TGCTGGGGG	GCCTTGGAGG	CCCTCTCTCC	CTGGACAGGG	TACCGTGGCT	23170
	TTTCTACTCT	GCTGGGCTG	CGGCTCGCG	TCAGGGCACC	AGCTCCGGAG	CACCCGCGGC	CCCATGTGTC	23240
30	ACGGAGTGCC	AGGCTGTGAG	CCACAGATGC	CCAGTCCAG	GTGTGGCCGC	TCCAGCCCCC	GTGCCCCCAT	23310
	GGGTGTTTTT	GGGGGAAAAG	GCCAGGGGCA	GAGGTGTGAG	GAGACTGGTG	GGCTCATGAG	AGCTGATTCT	23380
	GCTCCTTGCC	TGAGCTGCCC	TGAGCAGCCT	CTCCCGCCCT	CTCCATCTGA	AGGGATGTGG	CTCTTCTTAC	23450
	CTGGGGCTCC	TGCCCTGGGG	CAGCCTTGGG	CTACCCCACT	GGCTGTACCA	GAGGGACAGG	CATCCTGTGT	23520
	GGAGGGGCAT	GGGTTTCAGT	GGCCCCAGAT	GCAGCCTGGG	ACCAGGCTCC	CTGGTGCTGA	TGGTGGGACA	23590
35	GTCAACCTGG	GGGTTGACCG	CGGACTGGG	CGTCCCAAGG	GTTGACTATA	GGACAGGTG	TCCAGGTGTC	23660
	CTGCAAGTAG	AGGGGCTCTC	AGAGGCGTCT	GGCTGTCATG	GGTGGACGTG	GCCCCGGGCA	TGGCCTTCAG	23730
	CGTGTGCTGC	CGTGGGTGCC	CTGAGCCCTC	ACTGAGTCTG	TGGGGGCTTG	TGGCTTCCCG	TGAGCTTCCC	23800
	CCTAGTCTGT	TGTCTGGCTG	AGCAAGCCTC	CTGAGGGGCT	CTCTATTGCA	GACAGCACTT	TGACCAATAT	23870
	CAGCTGCGGG	AGCTGTGCGA	AGCAGAGGTC	AGGCAGCATC	GGGAAGCCAG	GCCCCGCCCT	CTGAGCTCCA	23940
40	GACTCTCGCT	CATCCCCAAG	CTGACGGGG	TGCGGCCGAT	TGTGAACATG	GACTACGTCT	TGGGAGGGTG	24010
	AACGTTCGCG	AGAGAAAAGA	GGGTGGCTGT	GCTTGGTTT	AACCTCTCTT	TTAAACAGAA	GTGCGTTTGA	24080
	CCCCACATT	TGGTATCAGC	TTAGATGAAG	GGCCCGGAGG	AGGGGCCACG	GGACACAGCC	AGGGCCATGG	24150
	CACGGCGCCA	ACCCATTGTT	CGCACAGTGT	AGGTGGCCGA	GGTGCCGGTG	CCTCCAGAAA	AGCAGGCTGG	24220
	GGGTGTAGGG	GGAGCTCTCT	GGGCAGGGAG	AGGCTCTGAG	GACCACAAGA	AGCAGCCGGG	CCAGGGCCTG	24290
45	GATGACGAC	GGCCCGAGGT	CTGGATCCG	TGTCTGCTG	TGGTGCCGAG	CCTCCGTGCG	CTTCCGCTTA	24360
	CGGGGGCCGG	GGACCAAGCC	AGCACTGCCA	GGAGCCCAAC	GGGCTCTGAG	GATCCTGGAC	CTTGGCCCCA	24430
	GGCTCTGCA	CCCCACCCCT	TGGCTGCGG	TGGCTGCGGT	GACCCCGTCA	TCTGAGGAGA	GTGTGGGGTG	24500
	AGGTGGACAG	AGGTGTGGCA	TGAGGATCCC	GTGTGCAACA	CACATGCGGC	CAGGAACCCG	TTTCAAACAG	24570
	GGTCTGAGGA	AGCTGGGAGG	GGTCTAGGTT	CCCGGCTCTG	GGTGGCTGGG	GACACTGGGG	AGGGGCTGCT	24640
50	GTCTCCCTGG	GTCCCTATGG	TGGGGTGGGC	ACTTGGCCGG	ATCCACTTTC	CTGACTGTCT	CCCATGCTGT	24710
	CCCCGCCAGG	CCGAGCGTCT	CACCTCGAGG	GTGAAGGCAC	TGTTACAGCGT	GCTCAACTAC	GAGCGGGGCG	24780
	GGCGCCCGGG	CCTCCTGGGC	GCCTCTGTGC	TGGGCTTGA	CGATATCCAC	AGGGCCTGGC	GCACCTTCGT	24850
	GCTGCGTGTG	CGGGCCCAAG	ACCCGCCCGC	TGAGCTGTAC	TTTGTCAAGG	TGGGTGCGGG	GGACCCCGCT	24920
	GAGCAGCCCT	GCTGGACCTT	GGGAGTGGCT	CCCTGATTGG	CACCTCATGT	TGGGTGGAGG	AGGTACTCCT	24990
55	GGGTGGGCGG	CAGGGAGTGC	AGGTGACCTC	GTCACTGTTG	AGGACACACC	TGGCACCTAG	GGTGGAGGCC	25060
	TTACAGCTTT	CCTGCAGCAC	ATGGGGCCGA	CTGTGCAACC	TGACTGCCCC	GGCTCCTATT	CCCAAGGAGG	25130
	GTCCCACTGG	ATTCCAGTTT	CCGTGAGAGA	AGGAACCGCA	ACGGCTCAGC	CACCAAGGCC	CGGTGCCCTG	25200
	CACCCAGTCT	CTGAGCCAGG	GGTCTCCTGT	CCTGAGGCTC	AGAGAGGGGA	CACAGCCCGC	CCTGCCCTTG	25270
	GGGTCTGGAG	TGGTGGGGGT	CAGAGAGAGA	GTGGGGGACA	CCGCCAGGCC	AGGCCCTGAG	GGCAGAGGTT	25340
60	ATGTCTGAGT	TTCTGCGTGG	CCACTGTGAG	TCTCTCTGCC	TCCACTCACA	CAGGTGGATG	TGACGGGGCG	25410
	GTACGACACC	ATCCCCCAGG	ACAGGCTCAC	GGAGGTCATC	GCCAGCATCA	TCAAACCCCA	GAACAGGTAC	25480
	TGCGTGGCTC	GGTATGCCGT	GGTCCAGAAG	GCCGCCCATG	GGCACGTCCG	CAAGGCCTTC	AAGAGCCACG	25550
	TAAGGTTTAC	GTGTGATAGT	CGTGTCCAGG	ATGTGTGTCT	CTGGGATATG	AATGTGTCTA	GAATGCACTG	25620
	GTGTCTGTGA	TGCGTTTCTG	TGGTGGAGGT	ACTTCCATGA	TTTACACATC	TGTGATATGC	GTGTGGGACA	25690
65	CGTGTGTGTC	GTGGTGCATG	TATCTGTGGC	GTGCATATTT	GTGGTGTGTG	TGTGTGTGGC	ACGTGTGTGT	25760
	CCATGGTGTG	TGTGCTGTGT	GTGTGATGTG	GTGTGTCTCT	GTGACACGTG	CATGTTTCATG	CTGTGTGCTG	25830
	CATGTCTGTG	ATGTGCCCTAT	TTGTGGTGTG	TGTGTGCATG	TGTCCGTGAC	ATATGCGTGT	CTATGGCATG	25900
	GGTGTGTGTG	GCCCCCTTGGC	CTTACTCCTT	CCTCCTCCAG	GCATGGTCCG	CACCATTTGT	CTCAGCTCTC	25970
	CGGGTGTCTG	TTTGGGGAGC	TCCACATTCA	GGGTCTCTAC	TCTTAGCATG	GGTGCCCTTG	TCCTGTCA	26040
70	GGGTGGGCCC	TTGGAGACTG	TAAGCCAGGT	TTGAGAGGAG	AGTAGGGATG	CTGGTGGTAC	CTTCTGGGAC	26110
	CCCTGGCACC	CCCAGGACCC	CAGTCTGGCC	TATGCCGGCT	CCATGAGATA	TAGGAAGGCT	GATTACGGCC	26180
	TCGTCCCCCG	GGACACACTC	CTCCAGAGGC	GGCCGGGGGC	CTTGGGGCTC	GGCAGGGGTG	AAAGGGGGCC	26250
	TGGGCTTGGG	TTCCACCCCA	GTGGTCATGA	GCACGCTGGA	GGGGTAAGCC	CTCAAAGTCG	TGCCAGGCGG	26320
	GGGTGCAGAG	GTGAAGAAGT	ATCCCTGGAG	CTTCGGTCTG	GGGAGAGGCA	CATGTGGAAA	CCCAACAGGA	26390
75	CCTCTTTCTC	TGACTTCTTG	AGCT					26414

Contig 2:

	TGTGGGATTG	GTTTTCATGT	GTGGGATAGG	TGGGGATCTG	TGGGATTGGT	TTTTATGAGT	GGGGTAACAC	70
	AGAGTTCAAG	GCGAGCTTTC	TTCCTGTAGT	GGGTCTGCAG	GTGCTCCAAC	AGCTTTATTG	AGGAGACCAC	140
5	ATCTTCCTTT	GAACATATGG	CGGGTTTATA	GTAAGTCAGG	GGTGTGGAGG	CCTCCCTGGG	GCTCCCTGTT	210
	CTGTTTCTTC	CACCTCTGGG	TCGTGTGGTG	CCTGCTGTGG	TGTTGTGGCCG	GTGGGCAGGG	CTTCCAGGCC	280
	TCCTTGTGTT	CATTGGCCCTG	GATGTGGCCC	TGGCTACGCT	CCGTCTCTGG	AATTCCCTCG	CGAGTTGGAG	350
	GCTTCTTTTC	TTTCTTTTTT	TCCTTCTTTT	TTTTTTTTTT	TGATAACAGA	GTCTCGCTCT	TTTTTGCCCA	420
10	GGCTGGAGTG	GTTTGGCGTG	ATCTTGGCTC	ACTGCAACCT	GTGCTTCCTG	AGTTCAAGCA	ATTCTCTTGC	490
	CTCAGCCTCC	CAAGTAGCTG	GAATTATAGG	CGCCCAACAC	CATGCTGACT	AATTTTGTGA	ATTTAGTAG	560
	AGACGAGGTT	TCTCCATGTT	GGCCAGGCTG	GTCTCGAACT	CCTGACCTCA	GGTGATCCTC	CCACCTCGGC	630
	CTCCCAAAGT	GCTGGGATGA	CAGGTGTGAA	CCGCCGCCCC	CGGCCGAGAC	TCGCTTCCTG	CAGCTTCCTG	700
	GAGATCTGCA	GCGATAGCTG	CCTGCAGCCT	TGGTGCTGAC	AACCTCCGTT	TTCTTCTCC	AGGTCTCGCT	770
15	AGGGGTCTTT	CCATTTCATG	ACTCTCTTCA	CAGAAGAGTT	TCACGTGTGC	TGATTTCCCG	GCTGTTTCTT	840
	GGGTAAATTGG	TGTCTGCTGT	TTATCGATGG	CCTCCTTCCA	TTTCCTTTAG	GCTTTGTTTA	TGTTGTGTTT	910
	TCCCGCTCCT	TGAAGGAAAA	GTTTCGATTA	TGGATGTTTG	AACCTTCTTT	TCTAAACAAG	CATCTGAAGT	980
	TGCCGTTTTTC	CCTCTAAAGC	AGGGATCCCG	AGGCCCTGGG	CTGTGGAGTG	GCACCGGTCT	GGGGCTGTT	1050
	AGGAACCCGG	CGCACAGCGG	GAGGCTAGGT	GGGGTGTGGG	GAGCCAGCGT	TCCCCTCTGA	GGCCCGCCCC	1120
20	TCTCAGATCA	GCAGTGGCAT	GCGGTGCTCA	GAGGCGCACA	CACCTTACTG	AGAACTGTGC	GTGAGAGGGG	1190
	TCTAGATTCT	GTGCTCCTTA	TGGGAATCTA	ATGCCCTGATG	ATCTGAGGTG	GAACCGTTTG	CTTCCAAAAA	1260
	CATCCCTTTC	CCCACGTGCT	TCCTGTGGAA	AAATCGTCTT	CCACGAAACC	AGTCCCTGGT	ACCACAAATG	1330
	TTGGGGACCC	TGTGCTAAAG	ACCTGCTTCA	GCAGCCTCTC	GTCACTGTTG	ATATATTGGC	TTTCTGTGTT	1400
	TGAGTCCAGA	ATAATTACGG	ATTCTGTGTA	TGCTTTCGCG	CGACCTCAGA	CCCATGGGCT	ATTTGTGGGC	1470
25	GTGTTGCTCG	CTCCTGGGTT	GGGAAGGGTG	CAGGCCCAT	GTACCTTCCT	GTTACTGCCT	TCCAGGTGTT	1540
	TTCTCAGGGT	TGAATCGTAC	TCCATGTGGT	TTTAGCCAC	GGCCCTGCCG	CCAGCTCCTG	GGGGCTGGGG	1610
	AACATGCTGA	AGCACAGAGT	CACCGTGCGC	GTCTTTTGAT	GCCTCACAA	CTCGAGGCTT	CCTGTGTCGG	1680
	TGTTAGTGTG	TGTCACGTGC	CTGCTCACAT	CCTGTCTTGG	GGACGCAGGG	GCTTAGCAGG	TCCCGTAGTA	1750
	AATGACAAGC	GTCTTGGGGG	AGTCTGCAGA	ATAGGAGGTG	GGGGTGCCGG	TCTCTCTCCC	GGCTCTTCAG	1820
	ACTCTTCTCC	TGCCCTGTCT	GTGGCTGCAC	CTGCATCCCT	GCAATCCCTC	CAGCACTGGG	CTGGAGAGGC	1890
30	CCGGGAGCTC	GAGTGCCACT	TGTGCCACGT	GACTGTGGAT	GGCAGTCGGT	CACGGGGGTC	TGATGTGTGG	1960
	TGACTGTGGA	TGGCGGTTGG	TCACAGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCT	TGGGGTCTGA	2030
	TGTGGTGACT	GTGGATGGCG	GTCGTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCT	TGGGGTCTGA	2100
	TGTGGTGACT	GTGGATGGCG	GTCGTGGGGT	CTGATGTGGT	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	2170
35	TGGTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	2240
	TGTGGTGACT	GTGGATGGCG	GTCGTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCT	TGGGGTCTGA	2310
	TGTGTGGTGA	CTGTGGATGG	CGGTCTGGGG	GTCTGATGTG	GTGACTGTGG	ATGGCGGTCT	TGGGGTCTGA	2380
	TGTGTGGTGA	CTGTGGATGG	TGATCGGTCA	CAGGGGTCTG	ATGTGTGGTG	ACTGTGGATG	CGCGTCTGTT	2450
40	GGTCTGATGT	TGGTGACTGT	TGGATGGTGA	TCGGTCAACG	GGGTCTGATG	TGTGGTGACT	TGGATGGGCG	2520
	GTCTGTGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTGG	GTCCCGGGGG	TCTGATGTGT	GGTGACTGTG	2590
	GATGGCGATC	GGTCAACAGG	GTCTGATGTG	TGGTGACTGT	GGATGGCGGT	CGTGGGGTCT	GATGTGTGGT	2660
	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	GTCGTGGGGT	CTGATGTGGT	2730
	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	TGGTGACTGT	GGATGGCGGT	CGTGGGGTCT	GATGTGTGGT	2800
45	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	TGGTGACTGT	GGATGGCGGT	CGTGGGGTCT	GATGTGTGGT	2870
	GTGGTGACTG	TGGATGGCAG	TGCTGGGGGT	TGATGTGTGG	TGACTGTGGA	TGGCGGTCTG	GGGGTCTGAT	2940
	GTGTGGTGAC	TGTGGATGGC	GGTCTGGGGG	TCTGATGTGT	GGTGACTGTG	GATGGCGGTC	TGGGGTCTG	3010
	ATGTGTGGTG	ACTGTGGATG	GCGGTCTGTG	GGTCTGATGT	GGTGACTGTG	GATGGCGGTC	TGGGGTCTG	3080
	ATGTGTGGTG	ACTGTGGATG	GCGGTCTGTG	GGTCTGATGT	GGTGACTGTG	GATGGCGGTC	TGGGGTCTG	3150
50	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	GTCGTGGGGT	CTGATGTGGT	GACTGTGGAT	GGCGGTCTGT	3220
	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	GTCGTGGGGT	CTGATGTGGT	GACTGTGGAT	GGCGGTCTGT	3290
	GTACACAGGG	TCTGATGTGT	GGTGACTGTG	GATGGCGGTC	GTGGGGTCTG	ATGTGTGGTG	ACTGTGGATG	3360
	GCGGTCTGTT	GGTCTGATGT	GTGGTGACTG	TGGATGGCGG	TGCTGGGGTC	TGATGTGTGG	TGACTGTGGA	3430
	TGGCGGTCTG	GGGGTCTGAT	GTGGTGACTG	TGGATGGTGA	TGCGTCAACG	GGGTCTGATG	TGTGGTAGCT	3500
55	GCAGGTGGAG	TCCACAGTGT	GTCTGTAGCT	ACTTTGCGTC	CTCGCCCCC	CGGCCCTTCC	TCCCAAAACA	3570
	GAAGCTTCCC	AGGCGCTCTC	TGGGCTTCAT	CCCGCCATCG	GGCTTGGCCG	CAGGTCCACA	CGTCTGATC	3640
	GGAGAAACAA	AGTGGCCAGC	TCTGGCCGGG	GCAGGCCACA	TTTGTGGCTC	ATGCCCTCTC	CTCTGCCGGC	3710
	AGGTCTCTAC	CTTGACAGAC	CTCCAGCCGT	ACATGCGACA	GTTCTGGGCT	CACCTGCAGG	AGACAGAGCC	3780
	GCTGAGGGAT	CCCGTCTGTA	TCCAGCAGGT	CTGGGCACTG	CCCTGCAGGG	TTGGGCAACG	ACTCCAGACA	3850
60	GTGGGTCTCT	CCCTGGGCAA	TCACTGGGCT	CATGACCGGA	CAGACTGTTG	GGCTTGGGGG	GCAGTGGGGG	3920
	GAATGAGCTG	TGATGGGGGG	ATGATGAGCT	GTGTGCCCTG	GCGAAATCTG	AGCTGGGGCA	TGCCAGGCTG	3990
	CGACAGCTGC	TGCATTACAG	CACCTGCTCA	CGTTTGACTG	CGCGGCTCT	CTCCAGTTCC	GCAGTGCTTT	4060
	TGTTTATGAT	TTGCTAAATG	TCTTCTCTGC	CAGTTTTGAT	CTTGAGGCCA	AAGGAAAGGT	GTCCCTCTCC	4130
	TTTAGGAGGG	CAGGCCATGT	TTGAGCCGTG	TCCTGCCAG	CTGGCCCTTC	AGTGCTGGGT	CTGAGGCCAA	4200
65	AGGAAACGTG	TCCCTTCTCT	TAGGAGGACG	GGCCGTGTTT	GAGCCACGCC	CCGCTGAGCG	GGCTCTCTAG	4270
	TGCTGGGTCT	TCCACGTGTT	CCCTGTGGCC	CTTTGCAGAT	GTGGTCTGTC	CACGTGGGCC	TGTGGCTCTT	4340
	TGCAGATGCC	TGTTAGCACT	TGCTCGGCTC	TAGGGGACAG	TGCTGTCCAC	CGCATGAGGC	TCAGAGACCT	4410
	CTGGGCGAAT	TTCTTGGCT	CCCAGGGTGG	GGGTGGAGGT	GGCCTGGGCT	GCTGGGACCC	AGACCTGTG	4480
70	CCCGGACAGT	GGGCAGCAAC	TCTTGGATCA	CATATGCCAT	CCGGGCCACG	GTGGGTGTG	GTGGGTGTG	4550
	CCCGGACAGT	GGGCAGCAAC	TCTTGGATCA	CATATGCCAT	CCGGGCCACG	GTGGGTGTG	GTGGGTGTG	4620
	CTCGAGAGAC	TGGGCCCGGC	CAGCCCAACA	TGGCCCTGCA	TTCCAGCCCA	GCCCGGACCT	CTATCAAAA	4690
	CACTGACCCC	AAAAGGGACG	GAGGGTCTTG	GCCACGTGGT	CCTGCTGTGC	TCAGCACCCA	CCGGCTCACT	4760
	CCCATGTGTC	TCCCGTCTGC	TTTCGACAG	CTCCTCCCTG	AATGAGGCCA	GCAGTGGGCT	CTTCGACGTC	4830
	TTCTTACGCT	TCATGTGCCA	CCACGCCGTG	CGCATCAGGG	GCAAGTGAGT	CAGGTGGCCA	GGTGCCATTG	4900
	CCCTGCGGGT	GGCTGGGCGG	GCTGGCAGGG	CTTGTGCTCA	CCTCTCTCCT	GCCCTTCCC	CACTGNCCTT	4970

	CTGCCCGGGG	CCACCCAGAGT	CTCCTTTTCT	GGCCCCCGCC	CCCTCCGGCT	CCTGGGCTGC	AGGCTCCCGA	5110
	GGCCCCGGAA	ACATGGCTCG	GCTTGGCGCA	GGCGGAGCGG	AGCAGGTGCC	ACACGAGGCC	TGGAAATGGC	5180
	AAGCGGGGTG	TGGAGTTGCT	CCTGCGTGGA	GGACGAGGGG	CGGGGGGTGT	GTCTGGGTCA	GGTGTGCGGC	5250
5	GAGCGTTTGA	GCCTGCAGCT	TGTCAGCTCC	AAGTTACTAC	TGACGCTGGA	CACCCGGCTC	TCACACGCTT	5320
	GTATCTCTCT	CTCCCGATAC	AAAAGGATTT	TATCCGATTC	TCATTCTCTG	CCCTGTCTGT	TGACCCCGCG	5390
	GAGGGCGCGG	GCTCTTCTCT	CTGTGACTAG	ATTTCCCATC	TGGAAAGTGC	GGGGTTGACC	GTGTAGTTTG	5460
	CTCCCTCTCG	GGGGCCCTGT	GTGGCCATGG	GGCAGGCGGC	CTGGGAGAGC	TGCCGTGACA	CAGCCACTGG	5530
	GTGAGCCACA	CTCACGGTGG	TAGAGCCACA	GTGCTTGGTG	CCACATCAGC	TCCTCTGGAT	TTTAAGTAAA	5600
	ACCACACACC	TCCCGGCAGG	CATCTGCCTG	CGACCCCTGT	TGTGCTGGG	GAGAGTGGTA	GCACGGAGGA	5670
10	AATTCGTGCA	CACCTCAAGG	CATCAGCAAG	GTCCATCCGA	GTGAGGTGGA	ACGTGGAGGC	CTCTCTCTGG	5740
	GATCGTCTCC	AGCGGATAAA	GGACTGTGCA	CAGCTTCGGA	AGCTTTTATT	TAAAAATATA	ACTATTAATT	5810
	ATTGCAATTAT	AAGTAATCAC	TAATGGTATC	AGCAATTATA	ATATTTATTA	AAGTATAATT	AGAAATATTA	5880
	AGTAGTACAC	ACGTTCTGGA	AAAACACAAA	TTGCACATGG	CAGCAGAGTG	AATTTTGGCC	GAGGGACACG	5950
	TGTGCACATG	TGTGTAAGCG	GGCCCCAGGC	CCACAGAAAT	CGCTGACAAA	GTCACCTCCC	CAGAGAAGCC	6020
15	ACCACGGGCG	TCCTTCGTGG	TCGTGAATTT	TATTAAGATG	GATCAAGTCA	CGTACCGTCC	ACGTGTGGCA	6090
	GGGCTTTGGG	GAATGTGAGG	GATGACTGTC	GTCCCTCATG	CCTGACAGAC	AGGAGGTGAC	TGTGTCTGTC	6160
	CTGTCCCTAG	GACACGGACA	GGCCCGAAGC	TCTAGTCCCC	ATCGTGGTCC	AGTTTGGCCT	CTGAATAAAA	6230
	ACGTCTTCAA	AACCTGTTGC	CCCAAAACT	AAGAACAGAG	AGAGTTTCCC	ATCCCATGTG	CTCAGAGGGG	6300
	CGTATCTGCT	TGCGTTGACT	CGCTGGGCTG	GGCCGACTCC	TAGAGTTGGT	CGGTGTGCTT	CTGTGCAAAA	6370
20	AGTGCACTCC	TCTTGCCCAT	CACTGTGATA	CTGCAACAG	CAAGGAAAGC	CTCTTTTCTT	TTCTTTCTTT	6440
	TTTTTTTTTT	GAGACGGAAC	GTCACTGTGG	TCTGCTGGG	CTTGAGTGCA	GTGGCGCGAT	CTCAACTCAC	6510
	TGCAACCTCC	GCCTCCCGGG	TTCCAGCATT	TCTCTGCTCT	CAGCCTCCCC	AGCAGCTGAG	ATTACAGGCA	6580
	CCCACCCCTC	GGCGCTGGCT	AATTTTGTGA	TTTTTAGTAG	AGAGGGGTTT	TGCCCATGTT	GGCCAGGCTG	6650
	GTCTCGAACT	CCTGACCTCA	GGTGATCCAC	CCACCTCGGC	CTCCCAAGT	GCTGGGATTA	CAGGTGTGAG	6720
25	CCATCACGCC	CAGCCGGAAG	GCCTCTTTTT	AAGGTGACCA	CCTATAGCGC	TTCCGAAAA	TAACAGGCTG	6790
	TGTTTTTGCA	GTAGGCTGCA	AGCCTCTCTT	AGCAACAGGA	GTGGCTCTCT	GTGGGCTCTG	GGGATGGCTG	6860
	AGGGTCGCGT	GGCAGCCATG	CCTTCTGTGT	GCACCTTTAG	GTTCACCGGG	GCTATTCCTG	TCTCACTGTT	6930
	TGTCTGAAAA	CGCACCCCTG	GCATCCCTTG	TTGGAGAGTT	TCTGCTTCTC	GTTGGTCATG	TGAAACTGAG	7000
30	GGGCAAGGTT	GTATCCGTTG	GGCGCGAGCG	GCTACATGTA	GGGTCTAGAG	TCTTTCACCG	TGGACAAATT	7070
	CCTTGAAAAA	AAAAAAGGA	GTCCGGTTAA	GCATTCATTG	CGGGTCAAGT	GTCTGGTTCT	GTGAATAAAC	7140
	CTTAAGATTT	AAGAAACCTT	AATGAAAGAA	AACCTTGATG	ATTACAGACA	AGGATGTGGT	CACACCTGTG	7210
	GCTGGATCTG	TTTCAGCGCG	CCCAGTGCA	GGTGAGAGTG	GGGAGCAGGG	ATTGTTTGGT	CAGAGGCTCT	7280
	ATCTGGTATG	TTTCTGAGGT	GTTCGCCGCG	TGAATGGTAG	ACGTGTCGTT	TGTGTGTATG	AGGTTCTGTG	7350
35	TCGTGTGTGT	GCTCGGTTTG	AGTGTACGCA	TGTCCAGCAC	ATGCCCTGCC	CGTCTCTCAC	CTGTGCTTTC	7420
	CCGCCCCAGG	TCCTACGTCC	AGTGCCAGGG	GATCCCGCAG	GGCTCCATCC	TCTCCACGCT	GCTCTGCGAG	7490
	CTGTGCTACG	GCGACATGGA	GAACAAGCTG	TTTGCGGGGA	TTGCGCGGGA	CGGGTGAGGC	CTCCTCTTCC	7560
	CCAGGGGGGG	TTGGGTGGGG	GTGATTTGCG	TTTTGATGCA	TTCAAGTTGA	ATATTCCTGG	TGCTCTGGAG	7630
	ACCATGACTG	CTCTGTCTTG	AGGAACAGCA	CAAGGTTGCA	GGCCTTCTTT	GGTATGAAGC	CGCACGGGAG	7700
40	GGGTTGCACA	GCCTGAGGAC	TGCGGGCTCC	ACGCAGGCTC	TGTCCAGCGG	CCATGTCCAG	AGGCCCTCAG	7770
	GCTCAGCAGG	CGGGAGGGCC	GCTGCCCTGC	ATGATGAGCA	TGTGAATTCA	ACACCGAGGA	AGCACACGAG	7840
	CTTCTGTACG	GTACCCACAG	TTCCGTTAGG	GTCCCTGGGG	AGATGGGGCT	GGTGACAGCT	GAGGCCCCAC	7910
	ATCTCCGAGC	AGGCCCTCGA	CAGGTGGCCT	GGACTGGGCG	CCTCTTCAGC	CCATTGCCCC	TCCCACCTTG	7980
	ATGGGGTCTA	CACCCAGGA	CGCACACACC	TAAATATCGT	GCCAACTTAA	TGTGGTTCAA	CTCAGCTGGC	8050
45	TTTTATTGAC	AGCAATTACT	TTTTTTTTTT	TAATACTTTA	AGTTCTAGTG	TACATGTGCA	CGACGTGGAG	8120
	GTTAGTTACA	TATGTATACA	TGTGCCATGT	TGGTGTGCTG	CACCCATTAA	CTCATCATTT	ACATTAGGTA	8190
	TATCTCTTAA	TGCTATCCCT	CCCCACTCCC	CCCATCCCAT	GACAGGCCCT	GGTGTGTGAT	GTTCCTCCAC	8260
	CTGTGTCCAA	GTGTTCTCAT	TGTTCTGATT	CCACCTGTGA	GTGAGAACAT	GTGGTGTGTT	GTCTTCTTTC	8330
	CTTGCAATAG	TTTGCTCAGA	GTGATGGTTT	CCAGCTTCGT	CCATGTCCCT	ACAAAGGACA	TGAATCATAT	8400
50	CTTTTATTAT	ACTGCATAGT	ATTCGTTGGT	GTATATGTGC	CACATTTTCT	TAATCCAGTC	TATCATCGAT	8470
	GGACATTTGG	GTTGGTTGCA	AGTCTTTGCT	ACTGTGAATA	GTGCCGCAAT	AAACATACGT	GTGCATGTGT	8540
	CTTTATAGCA	GCATGATTTA	TAATCCTTTG	GGTATATACC	CAGTAATGGG	ATGGCTGGGT	CAAAATGGTAT	8610
	TTCTAGTTCT	AGATCCTTGA	GGAATCACCA	CACGTGCTTC	CACAATGGTT	GAACATAGTT	ACATCCCCAC	8680
	CAACAGTGTA	AAAGTGTCTT	GGTGTGGAG	AGGATGTGGA	CAGCAGTTAT	TTTTTTATGA	AAATAGTATC	8750
55	ACTGAACAAG	CAGACAGTTA	GTGAAGGATG	CGTCAGGAAG	CCTGCAGGCC	ACACAGCCAT	TTCTCTCGAA	8820
	GACTCCGGGT	TTTTCTCTGT	CATCTTTTGA	AACTCTAGCT	CCAATTATAG	CATGTACAGT	GGATCAAGGT	8890
	TCTTCTTCAT	TAAGGTTCAA	GTCTAGATT	GAAATAAGTT	TATGTAACAG	AAACAAAAAT	TTCTGTATCA	8960
	CACAACCTGC	TCTGGGATTT	GGAGGAAAGT	GTCCCTGAGC	TGGCGGCACA	CTGGTCAGCC	CTCTGGGACA	9030
	GGATACCTCT	GGCCCATGGT	CATGGGGCGC	TGGGCTTGGG	CCTGAGGGTC	ACACAGTGCA	CCATGCCGAC	9100
60	CTTCTGTGG	ATAGGATCTG	GGTCTCGGAT	CATGCTGAGG	ACCACAGCTG	CCATGCTGGT	AAAGGGCACC	9170
	ACGTGGCTCA	GAGGGGGCGA	GGTTCACAGC	CCCAGCTTTC	TTACCGTCTT	CAGTTATTTT	TCCCTAAGAG	9240
	TCGTGAGAA	GGGGCGCGGC	CTGATGGCCT	TCGTTCTGCT	TCAGCTGGCA	CAGAATTGCA	CAGCTGTATG	9310
	GTAAACACTG	AGTACTTATA	ATGAATGAGG	AATTGCTGTA	GCAGTTAACT	GTAGAGAGCT	CGTCTGTTGG	9380
	AAAGAAATTT	AAGTTTTTCA	TTTAACCGCT	TTGGAGAATG	TTACTTTTAT	TATGGCTGTG	TAAATGTTT	9450
65	GACATTCACT	CCCTCGTAGA	CAGATACTAC	GTAAAAAGTG	TAAAGTTAAC	CTTGCTGTGT	ATTTTCCCTT	9520
	ATTTTAGGCT	GCTCCTGCGT	TTGGTGGATG	ATTTCTTGTT	GGTGACACCT	CACCTCACCC	ACGCGAAAAA	9590
	CTTCTTCAGG	TGAGGCCCCG	GCCGTGTGTC	TGTGGGGACC	TCCACAGCCT	GTGGGCTTTG	CAGTTGAGCC	9660
	CCCCGTGTCC	TGCCCTGTGG	ACCGCAGCGT	TGTCTCTGCC	AAGTCTCTCT	TCTCTGCCGG	TGCTGGATCC	9730
	GCAAGAGCAG	AGGCGCTTGG	CCGTGCACCC	AGGCTTGGGG	GCGCAGGGGC	ACCTTCGGGA	GGGAGTGGGT	9800
70	ACCGTGCAGG	CCCTGGTCCT	GCAGAGACGC	ACCCAGGTTA	CACACGTGGT	GAGTGCAGGC	GGTGACCTGG	9870
	CTCCTGCTGC	TCTTTGGAAA	GTCAAGAGTG	GCGGCTCTCT	GGGCCCCAGT	GAGACCCCCA	GGAGCTGTGC	9940
	ACAGGGCCCTG	CAGGGCCGAG	GGGGAGCCTC	CTTCCCGAGG	GTGCACCTGA	GCCTGCGGAG	AGCAGGAGCT	10010
	GCTGAGTGAG	CTGGCCACAG	CGGTTCCGCT	CGGTCACTGT	CCTGCGTGGG	GTTGTTTGGG	ATCGGTGGGA	10080
	GAATTTGGAT	TTGCTGAGTG	GTGCTGTCTT	GAAACACGGA	GATGGCTAGG	AGTGGGTTTC	AGAGTTGATT	10150
75	TTTGTGAATC	AAACTAAAAT	CAGGCACAGG	GGACCTGGCC	TCAGCACAGG	GGATGTGCTA	ATGTGGTCTC	10220
	CCTCAAGGGC	GCCCCACAGA	CGCGGTGGGC	TGTTTTTAAA	GTGCGATTTC	ACGAGGGAGC	AGAAACCTTG	10290
	AAAGCTGTAA	AGGGAACCTT	CAGAAAATGT	GGCCGCCAGG	GGTGGTTTCA	GGTGCTTTGC	TGGGCTGTGT	10360
	TTGTGAAAAA	CCATTGAGAC	CGGCCCTCCA	AGTCCACCTT	CCAGGTCCAC	CCTCCAGGGC	CGCCCTGGGC	10430

	TGGGGGTATG	CCTGGCGTTC	CTTGTGCCGC	AGCCCCGAGC	ACAGCAGGCT	GTGCACATTT	AAATCCACTA	10500
	AGATTCACTC	GGGGGGAGCC	CAGGTCCCAA	GCAACTGAGG	GCTCAGGAGT	CCTGAGGCTG	CTGAGGGGAC	10570
	AGAGCAGACG	GGGAACGCTG	CTTCTGTGTG	GCAAGTTCCT	GAGGGTGCTG	GCCAGGGAGG	TGGCTCAGAG	10640
	TGTATGTGGG	GGTCCCACCG	GGGGCAGAAC	TCTGTCTCTG	ATGAGTCGGC	AGCCATGTAA	CAGGAAGGGG	10710
5	TGGCCACAGG	GAGCTGGGAA	TGCACCAGGG	GAGCTGCCGA	GCTGGCCGAG	GTCCCAGGGC	CAGGCCACAG	10780
	GAAGGGCAGG	GGGACGCCCC	GGCCACACGC	AGAGGCCGCA	GGAGGGGAAG	GGGATGCCCA	GGCCAGAGCA	10850
	GAGGCTACCG	GGCACAGGGG	GGCTCCCTGA	GCTGGGTGAG	CGAGGCTCAT	GACTCGGCCA	GGGAACCTCC	10920
	TTGAGCTGAA	GCTGACGACT	GGTGTGTGCC	AGCTCACAGC	CCAGCCAGGT	CCCGCGCCTG	AGCAGGAACT	10990
	CAGAACCCTC	CCCTTTGTCT	AAAGCACAGC	AGATGCCTTC	AGGGCATCTA	GGAGAAAACA	GGCAAAAGTC	11060
10	TTGAGAAACG	TCTTAAAGA	AGGTGGGATG	GTGGCAATTT	CTTGTCCAGA	TTTTAGTCTG	CCCCGGACCA	11130
	CAGATGAGTC	TATAACGGGA	TTGTGGTGT	GCCATGGGGA	CACATGAGAT	GGACCATCAC	AGAGGCCACT	11200
	GGGGCTGCAC	CTCCCATCTG	AGTCTGGGCT	GTCCCGGGTC	CAGGCCAGGT	TCTTGCATGC	TCACCTACCT	11270
	GTCCTGCCCG	GGAGACAGGG	AAAGCACCCC	GAAGTCTGGA	GCAGGGCTGG	GTCCAGGCTC	CTCAGAGCTC	11340
	CTGCCAGGCC	CAGCACCCCTG	CTCCAAATCA	CCACTTCTCT	GGGGTTTTCC	AAAGCATTTA	ACAAGGGTGT	11410
15	CAGGTTACCT	CCTGGGTGAC	GGCCCCGCAT	CCTGGGGCTG	ACATTGCCCC	TCTGCCTTAG	GACCTGTGTC	11480
	CGAGGTGTCC	CTGAGTAGTG	CTGCGTGGTG	AACTTGGCGA	AGACAGTGGT	GAACCTCCCT	GTAGAAGACG	11550
	AGGCCCTGGG	TGGCACGGCT	TTTGTTCAGA	TGCCGGCCCA	CGGCCTATTC	CCCTGGTGGG	GGCTGCTGCT	11620
	GGATACCCGG	ACCTCTGGAG	TGCAGAGCGA	CTACTCCAGG	TGAGCGCACC	TGGCGCGAAG	TGGAGCCTGT	11690
	GGCCGGCTGG	GGCAGGTGCT	GCTGCAGGGC	CGTTCGCTCC	ACCTCTGCTT	CCGTGTGGGG	CAGGGCAGCTG	11760
20	CCAATCCCAA	AGGGTCAGAG	GGCCACAGGT	GCCCTCTGCT	CCATCTGGGG	CTGAGCAGAA	ATGCATCTTT	11830
	CTGTGGGAGT	GAGGGTGCTC	ACAACGGGAG	CAGTTTTCTG	TGCTATTTTG	GTAAAAGGAA	ATGGTGACCC	11900
	AGACCTGGGT	GCACCTGAGT	GTCTTCAGAA	AGCAGTCTGG	ATCCGAACCC	AAGACGCCCG	GGCCCTGCTG	11970
	GGCGTGAGTC	TCTCAAACCC	GAACACAGGG	GGCCTGCTGG	GCATGAGTCC	CTCTGAACCC	GAGACCCCTGG	12040
25	GGCCCTGCTG	GGCGTGAGTC	TCTCCGAACC	CAGAGACTTC	AGGGCCCTTT	TGGGCGTGAG	TCTCTCCGCT	12110
	GTGAGCCCCA	CACCTCCAAGG	CTCATCCACA	GTCTACAGGA	TGCCATGAGT	TCATGATCAC	GTGTGACCCA	12180
	TCAGGGGACA	GGGCCATGGT	GTGGGGGGGG	TCTCTACAAA	ATTCTGGGGT	CTTGTTCCTC	CAGAGCCCCA	12250
	GAGCTCAAGG	CCCCGTCTCA	GGCTCAGACA	CAAATGAATT	GAAGATGGAC	ACAGATGCAG	AAATCTGTGC	12320
	TGTTTTCTTT	ATGAATAAAA	AGTATCAACA	TTCCAGGCAG	GGCAAGGTGG	CTCACACCTA	TAATCCACGC	12390
30	ACTTTGGGAG	GGCGAGGTGG	GTGGATCACT	TGAGGCCAGG	AGTTTGAGGC	CAACCTAACC	AAATAGTGA	12460
	AATTCCATTT	CTACTTAAAA	AATACAAAAA	TGAGCTGGC	CTGGTGGCAC	ACGCCTGTAG	TCCCCGCTAT	12530
	CGGGGAGGCT	GAGGCAGGAG	AATCATTGGA	ACCCAGGAGG	CAGAGGTTGC	AGTGAGCCGA	GATCACACCA	12600
	CTGCATCCCA	GGCTGGGCAA	CAGAGTGAGA	CTTCATCTTA	AAAAAAGTATC	AGCATTCCAA	12670	
	AACCATAGTG	GACAGGTGTT	TTTTTATTCT	GTCCCTCGAT	AAATATTACT	GGTGTGTGTC	TAGAGGCCGG	12740
35	AACCTGGGGT	GCCTTCTCT	GAAGGCACAC	CCTTCATGGG	AAGAGAAATA	AGTGGTGAAT	GGTGTGTAAT	12810
	CCAGAGGTTT	AAACTGGGGT	CCTGTCTGTC	TGAGTTAACA	GTCCAGATCT	GGACTTTGCC	TCTTTCCAGA	12880
	ATGCTCCCTG	GGGTTTGCTT	CATGGGGGAG	CAGCAGGTGT	GGACACCCTC	GTGATGGGGG	AGCAGCAGGT	12950
	GCAGAGCCCC	TCATGATGGG	GGAGTGGCAG	GTGCAGACAC	CCTTGTGCAT	GGTGCCACAG	ATGTCCCTGT	13020
	TGCAGCTCCC	TCCCCACAAG	GATGCCGGTC	TCTGTGCTC	CCACAGTCC	CTGCTTCCCT	CTCACAGGCT	13090
40	TACCTGGTCC	TGGCCTCCAC	TGGCTTTGTC	TGCATGATTT	CCACATTTC	TGGGCTCCCA	GCACCTCTTC	13160
	GCCTCTCCCA	GGCACCTCTG	CAGTGTCTGG	CATACCAGTC	AGCTGTGAAC	TGTCCACTGC	TTATTTTGCT	13230
	CCCATGAAA	TGTATTTTT	AGGACAGGCA	CCCTGGTTC	CAGCCTCTGG	CACAGCATCA	GTGAATGTTA	13300
	TTGAAGGACA	AAGGACAGAC	AAACAAATCA	GGAAATGGG	TTCTCTCTAA	ACACATTGCA	AAGCCACAGA	13370
	GGCTAGTGCA	GGATGGGTGG	GCATCAGGTC	ATCAGATGTG	GGTCCAATGC	CAGAAATATC	TGTGTCCCA	13440
45	AAGGCCACTT	GGTCAGAGTG	TGTGCTTGCA	GAGGTGGCTC	TAAAAGCTCA	GCAGTGGAGG	CAGTGGTTCC	13510
	CCATACTCAG	GGTGAATCTA	CATCTCTGT	GTCTGAAGTA	TACAGCAGAG	GCTTGAAGGG	CATCTGGGAG	13580
	AAGAAAACAG	GCAAAATGAT	TAAAGAAAGT	AAAAAAGGAA	AAGTGGTAAG	ATGGGAATTT	TCTTGTCCAG	13650
	ATTTTAGTCT	CCCAAAACAC	AGCTCAGATG	GTAGAATGTG	GTCAGAAGTC	ATGGACAGAA	CAATAGAAC	13720
	AAACGGGAAG	CCTATCTCTC	AGAAACGTGT	GTTAATGTGG	TATGTGGCAC	AGCTGATGGA	AAAGAGAGTG	13790
50	TGTGTGTAAT	TTTTTTTTCT	GAGAAAAGTC	ACTGGAAGCA	AATAAGTTGT	GTCTTTACAG	CATATACCAG	13860
	AGCAGATTCT	AGGTAGAAGA	GGAGACACAT	GCAAAACAA	CCAGCAACAG	AAATAAAACA	AAAGACTCAA	13930
	AGGGAAGGGA	GGTGAACGTT	CCCTGGTTTG	TGTGTGGGGA	AGGACACACA	GGGAGGCGGA	TGAACCCGCT	14000
	GAGGCAACGG	GCATTGCTTT	CACTGCAGAG	AAACTCAGCT	TGCTGAGCC	ACAGTGAATA	TGGCCATTCC	14070
	CTGGAGCGTT	TGTGCACGTG	ATTTATTTAA	GGGCCCTGT	GAGGTCTCTG	ACATTCTATCC	TCTCACTTTG	14140
55	TTCTCTTAAC	CACCTGAGAG	GTAGAGGAGG	AAAGGCTCCA	GGGGAGCAGC	CGCCCTTGGT	CACCCAGCTG	14210
	GCAAAAGGGA	TGCATGATTG	CAGCCTGGCC	TCTGTCTCCG	GGGCCCTTGC	TCTGCCCGAG	GACCCACAC	14280
	AAGTCAAGAC	CATAGGCTCA	GGGTGAGCCG	GAGCCCAAGG	TCGTGTTGGG	GATGGCTGTG	AAAGAAAGAA	14350
	TGGACGTCTG	ATGCACACTT	GGGAAGGTCC	TACCAGCAGC	GTCAAAGAAA	TGCATGTGAA	ACTGACAGCG	14420
	AGACCATCC	CTCAAAGAAA	CGCACGTGAA	ACTGATGGCG	AGACCTGTCC	CCATCCCTCA	TGCTGGCTCC	14490
60	TTTTCTGGGC	TTGCCAAGAG	CCAGCATCAG	GTTGAGGCAA	GCTGGAAGA	CTTTCTGGA	AAGCAGCTTG	14560
	TTTGATATGA	AGTCTCTACA	ATGTCTCTGT	TCTTCCAGT	AATTCACACT	CTGAAGTGAC	CAGACATTAT	14630
	CACGGGTCTT	ATTTACCATT	TCCAGTGTTC	CAGGCAGGGG	GACTTGCCAC	AGCAAGTCAC	GAACCTGCCC	14700
	AAATACAGGG	CTAAGGAGAT	ATTATGCATC	ACAAAACCTT	CTCTGCCATT	AAACATTTTT	CAAAAGATTT	14770
	TTGAAGAATG	TTTAATGGCA	CAAAACGTTT	ATTTCAATGT	AGCAGTGTTC	AAAGCTGGAT	GTAAAAGAAC	14840
65	ACACCCAGG	AGCCTGGCGT	GAATGTCTAT	TGTGTTTATC	TTTGACATG	GACATACATG	GGCAGTGAGT	14910
	GGTGGTGAGG	CCCTGGAGGA	CATCGGTGGG	ATGCCCTCAT	CCTGCCCTCT	TGGAGACACC	ATGTGTGCCA	14980
	CGTGCACTCA	CTGGAGCCCT	GTTTAGCTGG	TGCCACCTGG	CTCTTCCATC	CCTGAGATTC	AAACACAGTG	15050
	AGATTCCCCA	CGCCCAACTC	AGTGTCTTCC	CACAAAAAAC	CTGAGTCACA	CCTGTGTTCA	CTCGAGGGAC	15120
70	GCCCGGAGC	CAGGGCTCCA	CAGTTTATTA	TGTGTTTGTG	GCTGAGTTAT	GTGCAGATCT	CATCAGGGCA	15190
	GATGATGAGT	GCACAAACAC	GGCCGTGCGA	GGTTTGGATA	CACTCAACAT	CACTAGCCAG	GTCTTGGTGG	15260
	AGTTTGGTCA	TGCAGAGTCT	GGATGGCATG	TAGCATTGGG	AGTCCATGGA	GTGAGCACCC	AGCCCGCTCG	15330
	GGCTGCAGCG	CATGCCCCAG	GCAGGACAAG	GAAGCGGGAG	GAAGGCAGGA	GGCTCTTTGG	AGCAAGCTTT	15400
	GCAGGAGGGG	GCTGGGTGTG	GGGCAGGCAC	CTGTGTCTGA	CATTCCCCCC	TGTGTCTCAG	CTATGCCCGG	15470
	ACCTTCATCA	GAGCCAGTCT	CACCTTCAAC	GGCGGCTTCA	AGGCTGGGAG	GAACATGGGT	CGAAAGCTCT	15540
	TTGGGGTCTT	GCGGCTGAAG	TGTGCACGCC	TGTTTCTGGA	TTTGAGGGTG	AGCAGGCTGA	TGGTCAGCAC	15610
75	AGAGTTGACA	GTTTCAGGAG	TGTGTGCGCA	AGTATGTGTG	TGTGTGTGTG	CGCGGCTGCC	TGCAAGGCTG	15680
	ATGGTGACTG	GCTGCACGTA	AGAGTGCACA	TGTACGCATA	TACACGTGAG	CACATACATG	TGTGCATGTG	15750
	TGTACATGAA	GGCATGGCAG	TGTGTGCACA	GGTGTGCAAG	GGCAAGAGTG	TGTGCACATG	CGAATGCACA	15820

	CCTGACATGC	ATGTGTGTTT	GTGCACAGTC	GTGTGGGCAT	TCACGTGAGG	TGCATGCGTG	TGGGTGTGCA	15890
	GTGTGAGTAG	CATGTGTGCA	CATAACATGT	ATTGAGGGGT	CCTCGTGTTC	ACCCCGCTAG	GTCTCAGCA	15960
	CCAGTGGCC	TCCTTACAGG	ATGAGACGGG	GTCCAGGCC	TTGGTGGGCT	GAGGCTCTGA	AGCTCGAGCC	16030
	CTGAGGGCAT	TGTCCCATCT	GGGCATCCGC	GTCCACTCCC	TCTCTGTGG	GCTTCTGTGT	CCACTCCCC	16100
5	TCTCCTGTGG	GCATTTACAT	CCACTCCACT	CCCTCTCTCC	TGTGGGCATC	CGCGTCCACT	CCCCCTCTCT	16170
	GTGGGCATCT	GCGTCCACCT	CCCTCTCTGT	TGGGCATTTG	CGTCCACTCC	CTCTCCTGGT	TCCTTCTGT	16240
	CTTGGGCGAG	CCTCGGGGGC	AGGCAGATGA	CACAGAGTCT	TGACTCGCCC	AGGGTGGTTC	GCAGCTGCCG	16310
	GGTGAGGGCC	AGGCCGGATT	TCACTGGGAA	GAGGGATAGT	TTCTTGTCAA	AATGTTCCCT	TTTCTTGTTT	16380
	CATCTGAATG	GATGATAAAG	CAAAAAGTAA	AAACTTAAAA	TCCCAGAGAG	GTCTTCTACC	TTTCTCACTC	16450
10	TTTCTTGGCG	ACTCTAGGTG	AACAGCCTCC	AGACGGTGTG	CACCAACATC	TACAAGATCC	TCCTGCTGCA	16520
	GGCGTACAGG	TGAGCCGCCA	CCAAGGGGTG	CAGGCCACGC	CTCCAGGGAC	CCTCCGCGCT	CTGCTCACTT	16590
	GCTGACCCGG	GCTTCACCTT	GGAACTCTCT	GGTTTATAGG	GCAAGGAATG	TCTTACGTTT	TCAGTGGTGC	16660
	TGCTGCGCTG	GCACAGTTCT	GTTCCGGTGG	CTCTGTGCAA	AGCACCTGTT	CTCCATCTCT	GGGTAGTGGT	16730
	AGGAGCCGGT	GTGGCCCCAG	GTGCCCCAC	TGTGCCCTGT	CACTGGCCGT	GGGACGTCT	GGAGGCCATC	16800
15	CCAGGGCAGC	AGGGGCATGG	GGTAAAGAGA	TGTTTATGGG	GAGTCTTAGC	AGAGGAGGCT	GGGAAGGTGT	16870
	CTGAACAGTA	GATGGGAGAT	CAGATGCCCG	GAGGATTGGG	GGTCTCAGCA	AAGAGGGCCG	AGGTGGGTGC	16940
	AGGTGAGGGT	CGGTGGCCCC	ACCCCGGGGA	AGGTGCAGCA	GAGCTGTGGC	TCCCCACACA	CGCCCGCCAG	17010
	CACCTGTGCT	CTGGGCATGG	CTGTGCTCCT	GGAACTGTCC	CTGTCTGGG	TGGTCAGGGG	GTGCCCTCTG	17080
	CAAGAATCGA	CAACTTTATC	ACAGAGGGAA	GGGCCAATCT	GTGGAGGCCA	CAGGGCCAGC	TTCTGCTGGG	17150
20	AGTCAGGGCA	GGTGGTGGCA	CAAGCCTCGG	GGCTGTACCA	AAGGGCAGTC	GGGCACCACA	GGCCCGGGCC	17220
	TCCACCTCAA	CAGGCCTCCC	GAGCCACTGG	GAGCTGAATG	CCAGGAGGCC	GAAGCCCTCG	CCCCATGAGG	17290
	GCTGAGAAGG	AGTGTGAGCA	TTTGTGTTAC	CCAGGGCCGA	GGCTGCCGGA	ATTACCGTGC	ACACTTGATG	17360
	TGAAATGAGG	TCGTGCTCTA	TCGTGGAAAC	CCAGCAAGGG	CTCACGGGAG	AGTTTTCAT	TACAAGGTGC	17430
	TACCATGAAA	ATGGTTTTTA	ACCCGAGTGC	TTGCGCCTTC	ATGCTCTGGC	AGGGAGGGCA	GAGCCACAGC	17500
25	TGATGTTTAC	CGCCTTTGCA	CCAGCTCCAG	AGGCTTGGGA	CCAGGCTGTC	TCAGTTCAGC	GGTGGCTCCG	17570
	GCTCAGACCG	CCCTCCTCTC	TGCCTTCTCT	CTCTGCTCTA	AATCTTCCCT	GCTTTCGATC	TCCTTGACGC	17640
	GTGCTGGGG	CCTCGTGCAA	GCTGCTTGAC	TCCTTCCGG	AAACCTTTGG	GGTGTGCTGG	ATACAGGTGC	17710
	CACTGAGGAC	TGGAGGTGTC	TGACACTGTG	GTGAGCCCCA	GGGTCCAGCT	GGCGTGCTTG	GGGCCTCCTT	17780
	GGGCCATGAT	GAGGTTCAGAG	GAGTTTTCCC	AGGTGAAAAC	TCCTGGGAAA	CTCCAGGGGC	CATGTGACCT	17850
30	GCCACCTGCT	CCTCCCATAT	TCAGCTCAGT	CTTGTCTCTA	TTTCCCCACC	AGGGTCTCTA	GCTCCAGGGA	17920
	GCTCCCGTAG	AGGGCCTGGG	CTCAGGGCAG	GGCGGCTGAG	TTTCCCCACC	CATGTGGGGA	CCCTTGGGTA	17990
	GTGCTTTGAT	TGGGTAGCCC	TGAGGAGGCC	GAGATGCGAT	GGGCCACGGG	CCGTTTCCAA	ACACAGAGTC	18060
	AGGCACGTGG	AAGGCCCAGG	AATCCCTTTC	CCTCGAGGCA	GGAGTGGGAG	AACGGAGAGC	TGGGCCCCGA	18130
35	TTTACGGCCA	GCCAGGCTGC	AGTGGGCGAG	GCTGTGGTGG	TCCACGTGGC	GCTGGGGGGC	GGGTCTGATT	18200
	CAAAATCCGT	GGGGCTCGGC	CTTCTTGGCC	CGTGTGGGCC	CGCGCTCCAC	ACGGGCTTGG	GGTGGACGGC	18270
	CCGACCTCTA	GCAGGTGGCT	ATTCTCTCCT	TTGGAAGAGA	GCCCTCACC	CATGCTAGGT	GTCTCCCTCC	18340
	TGGGTGAGGA	GCGTGGCCGT	GTGGCAACCC	CGGGACCTTA	GGCTTATTTA	TTTGTTTAAA	AACATTCTGG	18410
	GCTTGGCTTC	CGTTGTTGCT	AAATGGGGAA	AAGACATCCC	ACCTCAGCAG	AGTTACTGAG	AGGCTGAAAC	18480
40	CGGGGTGCTG	GCTTGACTGG	TGTGATCTCA	GGTCAATCCA	GAAGTGGCTC	AGGAAGTCAG	TGAGACCAGG	18550
	TACATGGGGG	GCTCAGGCAG	TGGGTGAGAT	GAGGTACACG	GGGGGCTCAG	GCAGTGGGTG	AGGCCAGGTA	18620
	CATGGGGGGG	TCAGGCACATG	GGTGAGATGA	GGTACACGGG	GGGCTCAGGC	AGAGGGTCAG	ACCAGGTAGA	18690
	CGGGGGCTCT	GATCACACGC	ACATATGAGC	ACATGTGCAC	ATGTGCTGTT	TCATGGTAGC	CAGGCTCTGT	18760
	CACACCTGCC	CCAAAGTCCC	AGGAAGCTGA	GAGGCCAAAG	ATGGAGGCTG	ACAGGGCTGG	CGCGGTGGCT	18830
45	CACACCTGTA	GTCCAGCAGC	TTTGGGAGGC	CGAGGCGAGA	GGATCCCTTG	AGGCCAGGAG	TTTAAAGACA	18900
	GCCTGAGCAA	CATAGTAGAA	CCCCATCTCT	ATGAAAATA	AAAACAAAAA	TTAGCTGAAC	ATGGTGGTGT	18970
	GGCCCTGTAG	TTCCAATACT	TGGGAGGCTG	AAGTGGGAGG	ATCACTTGAG	CCCAGGAGGT	GGAGGTGCA	19040
	GTGAGCTGAG	ATTGCCACCA	TGTACTGTCAG	GTGGGTGAC	AGAGTGAGAG	CCCATCTCAA	CAACAACAAA	19110
	GAGAGCTGAC	AAATGCAGTT	TCTTGGAAAG	AAACATTTAG	TAGGAACTTA	ACCTACACAC	AGAAGCCAAG	19180
50	TCGGTGTCTC	GGGTGTCAGT	AGATGAGATG	ATGGGTCTCT	ACACCATCAC	CCCAGACCCA	GGGTTTATGC	19250
	ACCACAGGGG	CGGGTGGCTC	AGAAGGGATG	CGCAGGACGT	TGATATACGA	TGACATCAAG	GTGTGCTGAC	19320
	GAAGGGCAGG	ATTCATGATA	AGTACCTGCT	GGTACACAAG	GAACAATGGA	TAACTGGAA	ACCTTAGAGG	19390
	CCTTCCCGGA	ACAGGGGCTA	ATCAGAAGCC	AGCATGGGGG	GCTGGCATCC	AGGATGGAGC	TGCTTCAGCC	19460
	TCCACATGCG	TGTTCATACA	GATGGTGCAC	AGAAACGCAG	TGTACCTGTG	CACACACAGA	CACGCAGCTA	19530
55	CTCGCACACA	CAAGCACACA	CACAGACATG	CATGCATGCA	TCCGTGTGTG	TGCACCTGTG	CCCATGAGGA	19600
	AACCCATGCA	TGTGCATTCA	TGCACGCACA	CAGGCACCGG	TGGGCCCATG	CCCACACCCA	CGAGCACCGT	19670
	CTGATTAGGA	GGCCTTTCCCT	CTGACGCTGT	CGGCCATCCT	CTCAGGTTTC	ACGCATGTGT	GCTGCAGCTC	19740
	CCATTTTCATC	AGCAAGTTTG	GAAGAACCCC	ACATTTTTC	TGCGCGTCAT	CTCTGACACG	GCCTCCCTCT	19810
	GCTACTCCAT	CCTGAAAGCC	AAGAAGCGAG	GATGTGTCAG	GTGCTGGGCC	TCAGTGGCAG	CAGTGCCTGC	19880
60	CTGCTGGTGT	TAGTGTGTCA	GGAGACTGAG	TGAATCTGGG	CTTAGGAAGT	TCTTACCCCT	TTTCGCATCA	19950
	GGAAAGTGGT	TAACCCAAACC	ACTGTACGGC	TCGTCTGCC	GCCCTCTCCT	GGGGTGAGCA	GAGCACCTGA	20020
	TGGAAGGGAC	AGGAGCTGTG	TGGGAGCTGC	CATCCTTCCC	ACCTTGTCTT	GCCTGGGGAA	CGCTGGGGG	20090
	GCCTGGTCTC	TCCTGTTTGC	CCCATGGTGG	GATTGGGGG	GCCTGGGCTC	TCCTGTTTGC	CCTGTGGTGG	20160
	GATTGGGCTG	TCTCCGCTCC	ATGGCACTTA	GGGCCCTTGT	GCAAAACCAG	GCCAAGGGCT	TAGGAGGAGG	20230
65	CCAGGCCCGG	GCTACCCCGC	CCCTCTCAGG	AGCAGAGGCC	GCCTATCACC	ACGACAGAGC	CCCGCGCCGT	20300
	CCTCTGCTTC	CCAGTCACCG	TCCTCTGCC	CTGGACACTT	TGTCACGAT	CAGGGAGGTT	TCTGATCCGT	20370
	CTGAAATTCA	AGCCATGTGC	AACCTGCGGT	CCTGAGCTTA	ACAGCTTCTA	CTTTCTGTTC	TTTCTGTGTT	20440
	GTGGAATTTT	CACCTGGAGA	AGCCGAAGAA	AACATTTCTG	TCGTGACTCC	TGCGGTGCTT	GGGTCCGGAC	20510
	AGCCAGAGAT	GGAGCCACCC	CGCAGACCGT	CGGGTGTGGG	CAGCTTTCCG	GTGTCTCCTG	GGAGGGGAGC	20580
70	TGGGCTGGGC	CTGTGACTCC	TCAGCCTCTG	TTTTCCCCCA	GGGATGTCGC	TGGGGGCCAA	GGGCGCGCC	20650
	GGCCCTCTGC	CCTCCGAGGC	CGTGCACTGG	CTGTGCCACC	AAGCATTCCT	GCTCAAGCTG	ACTCGACACC	20720
	GTGTCACTTA	CGTGCCACTC	CTGGGGTTC	TCAGGACAGG	CAAGTGTGGG	TGGAGGCGAG	TGCGGGCCCC	20790
	ACCTTGCCCG	GGGTCTCTCT	TGAACGCCCT	GTGTGGGGCG	AGCAGCCTCA	GATGCTGCTG	AAGTGCAGAC	20860
	CGCCCGGGGC	CTGACCTTGG	GGGCTGGGAG	CCACGCTGGC	AGCCCTATGT	GATTAACGCG	TGGTGTCCCC	20930
	AGGCCACGGA	GCCTGGCAGG	GTCCCCAACT	TCTTGAACCC	CTGCTTCCCA	TCTCAGGGGC	GATGGCTCCC	21000
75	CACGCTTGGG	AGCCTTCTGA	CCCTGACCTT	GTGCTCTCTC	ACAGCCTCTT	CCCTGGCTGC	TGCCCTGAGC	21070
	TCCTGGGGTC	CTGAGCAAGT	TCTCTCCCGG	CCCCGCCGCT	CCAGCGTCAC	TGGGCTGCCT	GTCTGCTCGC	21140
	CCCGGTGGAG	GGGTGTCTGT	CCCTTCACTG	AGGTTCCACC	CAGCCAGGGC	CACGAGGTGC	AGGCCCTGCC	21210

5 TGCCCGGCCA CCCACACGTC CTAGGAGGGT TGGAGGATGC CACCTCTGGC CTCTTCTGGA ACGGAGTCTG 21280
 ATTTTGGCCC CGCAGCCAGC ACGCAGCTGA GTCCGGAAGCT CCCGGGGACG ACGCTGACTG CCCTGGAGGC 21350
 CGCAGCCAAC CCGGCACTGC CCTCAGACTT CAAGACCATC CTGGAGTGAT GGCCACCCGC CCACAGCCAG 21420
 GCGGAGAGCA GACACCAGCA GCCCTGTGAC GCCGGGCTCT ACGTCCGAGG GAGGGAGGGG CGGCCACAC 21490
 CCAGGCCCCG ACCGCTGGGA GTCTGAGGCC TGAGTGAGTG TTTGGCCGAG GCCTGCATGT CCGGCTGAAG 21560
 GCTGAGTGTC CGGCTGAGGC CTGAGCGAGT GTCCAGCCAA GGGCTGAGTG TCCAGCACAC CTGCGTCTT 21630
 CACTTCCCCA CAGGCTGGGC CTCGGCTCCA CCCCAGGGCC AGCTTTTCTT CACCAGGAGC CCGGCTTCCA 21700
 CTCCCCACAT AGGAATAGTC CATCCCCAGA TTCCGCCATT TCCACCCCTC GCCCTGGCCT CTTTGGCTT 21770
 10 CCACCCCCAC CATCCAGGTG GAGACCCCTGA GAAGGACCCCT GGGAGCTCTG GGAATTTGGA GTGACCAAG 21840
 GTGTGCCCCG TACACAGGCG AGGACCCCTGC ACCTGGATGG GGGTCCCTGT GGGTCAAATT GGGGGGAGGT 21910
 GCTGTGGGAG TAAATACTG AATATATGAG TTTTTCAGTT TTGAAAAA TCTCATGTTT GAATCTAAT 21980
 GTGCACTGCA TAGACACCAC TGTATGCAAT TACAGAAGCC TGTAGTGAA CGGGGTGGTG GTCACTGCGG 22050
 GCCCATGGCC TGGCTGTGCA TTTACGGAGG TCTATGAGTG AATGGGGTTG TGGTCAGTGC GGGCCCATGG 22120
 CCTGGCTGGG CCTGGGAGGT TTCTGATGCT GTGAGGCAGG AGGGGAAGGA GGGTAGGGGA TAGACAGTGG 22190
 15 GAGCCCCAC CCTGGAAGAC ATAACAGTAA GTCCAGGCCG GAAGGGCAGC AGGGATGCTG GGGGCCAGC 22260
 TTGGCGGCG GGGATGATGG AGGGCTGGC CAGGGTGCCA GGGATGATGG GGGCCGAGC TGGGTGGCA 22330
 GGGGTGATGG GGGGGCTGG TCTGGGTGGC GGGGAAGATG GGGAGCCCTG GCTGGGCCCC CTCTCCCTT 22400
 GCCTCCACC TGCAGCCGTG GATCCGGATG TGCTTCCCTG GTGCACATCC TCTGGGCCAT CAGCTTTTCT 22470
 GGAGGTGGGG GGCAGGGGCA TGACACCATC CTGTATAAAA TCCAGGATTC CTCCTCTCTG ACGCCCCAAC 22540
 20 TCAGGTGAA AGTCACATTC CGCCTCTGGC CATTCTCTTA AGAGTAGACC AGGATTCTGA TCTCTGAAAG 22610
 GTGGGTAGGG TGGGGCAGTG GAGGGGTGG ACACAGGAGG CTTACGGGTG GGGCTGGTGA TCTCTCTCA 22680
 TCCTCTTATC ATCTCCAGT CTCATCTCTC ATCCTCTTAT CATCTCCAG TCTCATCTGT TCTCTCTTA 22750
 TCTCCAGTC TCATCTGCA TCCTCTTACC ATCTCCAGT CTCATCTCTT ATCCTCTTAT CTCTAGTCT 22820
 25 CATCCAGACT TACCTCCAG GCGGGGTGCC AGGCTCGCAG TGGAGCTGGA CATACGTCTT TCTCAGGCA 22890
 GAAGGAACCT GAAGGATTGC AGAGAACAGG AGGGGCGGCT CAGAGGGAGC CAGCTTTGGG GTGAAGAAC 22960
 AGCCCCCTCT CAGAAGTTGG CTTGGGCCAC ACGAAACCGA GGGCCCTGCG TGAGTGGCTC CAGAGCCTTC 23030
 CAGCAGGTCC CTGGTGGGGC CTTATGGTAT GGCCGGGTCC TACTGAGTGC ACCTTGAGCA GGGCTTCTGG 23100
 TTTAGTGCA GCCCGGACGT GCCTGGTGTG GGGGTGGGG CTTATGGCCA CTGATATGCG GTCATTTAT 23170
 30 TGCTGCTGCT TCAGAGAATG TCTGAGTGAC CGAGCCTAAT GTGTATGGTG GGGCCAAGTC CACAGACTGT 23240
 GTCGTAAATG CACTCTGGTG CCTGGAGCCC CCGTATAGGA GCTGTGAGGA AGGAGGGGCT CTTGGCAGCC 23310
 GGCCTGGGGG CGCCTTTGCC CTGCAAACTG GAAGGGAGCG GCCCGGGCG CCGTGGGCGG ACGACCTCAA 23380
 GTGAGAGGTT GGACAGAAC GGGCGGGGAC TTCCAGGAG CAGAGGCCG TGCTCAGGCA CACCTGGGTT 23450
 TGAATCACAG ACCAACaGGT CAGGCCATTG TTCAGCTATC CATCTTCTAC AAAGCTCCAG ATTCCTGTTT 23520
 35 CTCCGGGTGT TTTTGTGTA AATTTTACTC AGGATTACTT ATATTTTGTG CTAAGTATT AGACCCCTAA 23590
 AAAAGGTATT TGCTTTGATA TGGCTTAACT CACTAAGCAC CTACTTTATT TGTCTGTTTT TATTTATTAT 23660
 TATTATTATT ATTAGAGATG GTGTCTACTC TGTCACCCAG GTTGTAGTG CAGTGGCACA GTCATGGCTC 23730
 GCTGTAGCCG CAAACCCCA GGCTCAAGTG ATCCTCCGGC CTCAGCTTCC CAGAGTGCTG GGATTACAGG 23800
 TGTGAGCCAC TGCCCTTTGCC TGGCACTTTT AAAAACCACT ATGTAAGGTC AGGTCCAGTG GCTTCCACAC 23870
 40 CTGTATCCCC AGTAGTTTGG GAAGCCGAGG CAGAAGGATT GTCTGAGGCC AGGAGTTTGA GACCAGCATG 23940
 GGTAAACATG GGAGACCCCA TCTCTACAAA AAATGCAAAA AGTTATCCGG GCGTGGGGTG CAGCATCTGT 24010
 AGTCCAGCT GCTCGGGAGG CTGAGTGGGA GGATCGCTTG AGCCCGGAG GTCATGGCTG CAGTGAGCTG 24080
 TGATTGTACC ATCGCACTCC AGCCTGGGCA ACAGAGTGAG ACCCTGTCTC AAAAAAAAAA AAAAAAAAAA 24150
 AAGGAGAAGG AGAAGAGAAG AAGAGGAAG AAGGAAAGAG AAGAAGAAG AAGAAGGAG 24220
 45 AAGGAGGCCT GCTAGGTGCT AGGTAGACTG TCAAACTCA GAGCAAAATG AAAATAACAA AGTTTTAAAG 24290
 GGAAGAAAA ACCCCAGCTC TTTGACTTTC CTTAGGCCGT GAACCTCATCT CAAGCAGCTT CCTTCCACAG 24360
 ACAAGCGTGT ATGGAGCGAG TGAGTTCAAA GCAGAAAGGG AGGAGAAGCA GGCAAGGGTG GAGGCTGTGG 24430
 GTGACACCAG CCAGGACCCC TGAAGGGAG TGGTTGTTTT CCGCTCAG CCCCACGCTC CTGCGGTGCC 24500
 TGCACCTGCT GTAACCGTCC ATGTTGGTGC CAGGTGCCCA CCTGGGAAGG ATGCTGTGCA GGGGGCTTGC 24570
 50 CAAACTTTGG TGGGTTTTCAG AAGCCCCAGG CACTTGTGGC AGGCACAATT ACAGCCCCCT CCAAGAGATG 24640
 CCCACGTCTT TCTCCTGGAA CTTGTGAATG TGTACCCGCG AAGGCAGAGG CTGGTGAAGG CTGACAGTGG 24710
 AATCAGCGCT GCCAGTCAGC CGATCTTAAG GTCATCTGG ATTATCTGGT GGGCCTGATA TGGCCACAAG 24780
 GGTCCCTAGA AGTGAGAGAG GGAGGAGGG GAGAGTCAGA GAGGGAGCT GAGAAGGACC ACTGGCCACT 24850
 55 GCTGGCTTTG AGATGGAGGA GGGGTCCCC AGCCAAGGAA TGGGGGCAGC CGCTCCATGC TGGAAAAGCA 24920
 AGCAATCTTC CCCGGTCTG AGGGCACAGC GCCCTGCCCA CGCCTCGATT TCAGGCCAGT GGGACCTGTT 24990
 TCAGCTTTCC GGCCTCCAGA GCTGTAAGAT GATGCGTTTG TGTTAGCCA CTAAGCTGCA GTGATTCGTC 25060
 ACAGCAGCAA ATGGAATAGC AGTACAGGGA AATGAATACA GGGACAGTTC TCAGAGTGAC TCTCAGCCCA 25130
 CCCCTGGG 25138

60 Beispiel 5

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz
 der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der
 Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist
 65 in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in
 5 den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon			Intron				3'-Exon			
Prä-mRNA	A/C	A	G		G	T	A/G	A	...	N C	A G G
Häufigk.(%)	70	60	80		100	100	95	70		80	100 100 60

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von
 15 ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

Intron 1: bp 11493-11596 (SEQ ID NO 4);

Intron 2: bp 12951-21566 (SEQ ID NO 5);

Intron 3: bp 21763-23851 (SEQ ID NO 6);

5 Intron 4: bp 24033-24719 (SEQ ID NO 7);

Intron 5: bp 24900-25393 (SEQ ID NO 8);

5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden

10 Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);

Intron 7: bp 3879-4858 (SEQ ID NO 11);

Intron 8: bp 4945-7429 (SEQ ID NO 12);

Intron 9: bp 7544-9527 (SEQ ID NO 13);

15 Intron 10: bp 9600-11470 (SEQ ID NO 14);

Intron 11: bp 11660-15460 (SEQ ID NO 15);

Intron 12: bp 15588-16467 (SEQ ID NO 16);

Intron 13: bp 16530-19715 (SEQ ID NO 17);

Intron 14: 19841-20621 (SEQ ID NO 18);

20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position
21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

Intron 1 (SEQ ID NO 4)

GTGGGCCTCCCCGGGGTCGGCGTCCGGCTGGGGTTGAGGGCGGCCGGGGGGAACAGCGACATGCGGAGAGCAGCGCAGG
CGACTCAGGGCGCTTCCCCCGCAG

5

Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCCGTGAGGGCCCAGGCCCCAGAGCTGAATGCAGTAGGGGCTCAGAAAAGGGGGCAGGCAGAGCC
CTGGTCCTCCTGTCTCCATCGTACGTGGGCACACGTGGCTTTTCGCTCAGGACGTGAGTGGACACGGTGATCTCTGCC
TCTGCTCTCCCTCCTGTCCAGTTTGCATAAACTTACGAGGTTACCTTCACGTTTTGATGGACACGCGGTTTCCAGGCGC
10 CGAGGCCAGAGCAGTGAACAGAGGAGGCTGGGCGCGGCAGTGGAGCCGGGTTGCCGGCAATGGGAGAAAGTGTCTGGAAG
CACAGACGCTCTGGCGAGGGTGCCTGCAGGTTACCTATAATCCTCTTCGCAATTTCAAGGGTGGGAATGAGAGGTGGGGA
CGAGAACCCCTCTTCTGGGGGTGGGAGGTAAGGGTTTTGCAGGTGCACGTGGTCAGCCAATATGCAGGTTTTGTGTTTA
AGATTTAATTGTGTGTTGACGGCCAGGTGCGGTGGCTCACGCCGTAATCCAGCACCTTTGGGAAGCTGAGGCAGGTGGA
TCACCTGAGGTGAGGAGTTTGAGACCAGCTGACCAACATGGTGAAACCTATCTGTACTAAAAATACAAAAATTAGCTG
15 GGCAATGGTGGTGTGTGCTGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGC
TGCAGTGAGCTGAGATTGTGCCATTGACTCCAGCCTGGGCGACAAGAGTGAAACTCTGTCTTTAAAAAAGAGTGT
CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGGAGATAAGACTGTTCTCCAGCACAGATCCTGGTCCCATCTTTAGGTAT
GAAGAGGGCCACATGGGAGCAGAGGACAGCAGATGGCTCCACCTGCTGAGGAAGGGACAGTGTGTTGGGTGTTTCAGGGG
ATGGTGCTGCTGGGCCCTGCGGTGTCCTCCACCTGTTTTCTGGATTGATGTTGAGGAACCTCCGCTCCAGCCCCCTTT
20 TGGCTCCAGTGCTCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCGATTTACCCCCCTCCCAAACTCCCAAGAC
ATGTAAGACTTCCGGCCATGCAGACAAGGAGGGTGACCTTCTTGGGGCTCTTTTTTTCTTTTTTTCTTTTTATGGTGGC
AAAAGTCATATAACATGAGATTGGCACTCCTAACACCGTTTTCTGTGTACAGTGCAGAATTGCTAACTCGGCGGTGTTTA
CAGCAGGTTGCTTGAAATGCTGCGTCTTGGGTGACTGGAAGTCCCTACCCATCGAACGGCAGCTGCCTCACACCTGCTGC
GGCTCAGGTGGACACGCGGAGTCAGATAAGCGTCATGCAACCCAGTTTTGCTTTTGTGCTCCAGCTTCCTTCGTTGAG
25 GAGAGTTTGAGTTCTCTGATCAGGACTCTGCTGTCTGCTGTTCTCTGACTTCAGATGAGGTGACAATCTGCCCCCTGG
CTTATGCAGGGAGTGAGGCGTGGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGTTGCCCCAGGTGTCCCT
GTCACGTGTAGGGTGAGTGAGGCGCGGCCCGGGTGTCCCTGTCCCGTGCAGCGTGATTGAGGTGTGGCCCCCGGGTGT
CCCTGTACGTGTAGGGTGAGTGAGGCGCCATCCCGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCGTGGTCCCCGG
GTGTCCCTGTCCCGTGCAGGGTGAGTGAGGCACTGTCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGCGGTCC
30 CCGGGTGTCCCTCTCAGGTGTAGGGTGAGTGAGGCGCGGCCCCAGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCACC
GTCCCTGGGTGTCCCTCCAGGTATAGGGTGAGTGAGGCACTGTCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGG
CGCGGCCCGGGTGTCCCTCTCAGGTGCAGGGTGAGTGAGGCGCTGTCCCTGGGTGTCCCTGTCTCGTGTAGGGTGAGT
GAGGCTCTGTCCCAGGTGTCTTGGCGTTTGCTCACTTGAGCTTGCTCCTGAATGTTTGCTCTTTCTATAGCCACAGCT
GCGCCGGTTGCCATTGCTGGGTAGATGGTGAGGCGCAGTGCTGGTCCCAAGCCTATCTTTCTGATGCTCGGCTCT
35 TCTTGGTCACTCTCCGTTCCATTTGCTACGGGGACACGGGACTGCAGGCTCTCGCCTCCCGGTGCCAGGCACTGCAG
CCACAGCTTCAGGTCCGCTTGCTCTGTTGGGCTGGCTGCTCACACGTGCCCGCCACATGCATGCTGCCAATACTCC
TCTCCAGCTTGTCTCATGCCGAGGCTGACTCTGGGCTGCTGTGCTGCTGCCAGTGTGCTGGAGACATCCAGAA
AGGGTTCTCTGTGCCCTGAAGGAAAGCAAGTCACCCCAGCCCCCTCACTTGTCTGTTTTCTCCCAAGCTGCCCTCTGC
TTGGCCCCCTTGGGTGGGTGGCAACGCTTGTACCTTATTCTGGGCACCTGCCGCTCATTGCTTAGGCTGGGCTCTGCTC
40 CCAGTCGCCCCCTCACATGGATTGACGTCCAGCCACAGGTGGAGTGTCTCTGTCTGTCTCCTGCTCTGAGACCCACGTG

GAGGGCCGGTGTCTCCGCCAGCCTTCGTGAGACTTCCCTCTTGGGTCTTAGTTTGAATTTCACTGATTACCTCTGACG
TTTCTATCTCTCCATTGTATGCTTTTCTTGGTTTATTCTTTCATTCCCTTTCTAGCTTCTTAGTTTAGTCATGCCTTTC
CCTCTAAGTGCTGCCTTACCTGCACCCTGTGTTTGTATGTGAAGTAATCTCAACATCAGCCACTTCAAGTGTCTTAAA
ATACTTCAAAGGTAAATACTTCTTTAAGTATTCTTATTCTGTGATTTTCTTGTGTCACGCTGTGTTTTCAGCTGA
5 AATCATTTTGATATCAGTGACTTTTAAGTATTCTTTAGCTTATTCTGTGATTTCTTTGAGCAGTGAGTTATTGAACACT
GTTTATGTTCAAGATATGTAGAGTATCAAGATACGTAGAGTATTTAAGTTATCATTTTATTATTGATTTCTAACTCAGT
TGTGTAGTGGTCTGTATAATACCAATTATTGAAGTTTGGGAGCCTTGCTTTGTGATCTAGTGTGTCATGGTTCCAG
AACTGTCCATTGTAAATTTGACATCCTGTCAATAGTGGGCATGCATGTTCACTATATCCAGCTTATTAAGGTCCAGTGCA
AAGCTTCTGTCTCCTTCTAGATGCATGAAATCCAAGAAGGAGGCCATAGTCCCTCACCTGGGGGATGGGTCTGTTTATT
10 TCTTCTCGTTTGGTAGCATTTATGTGAGGCATTGTTAGGTGCATGCACGTGGTAGAATTTTATCTTCTGATGAGTGAA
TCTTTTGGAGACTTCTATGTCTCTAGTAATCTAGTAATCTTTTAAATTTGCTCTTAGTACTGCCACACTGGGCTTCT
TTTGATTAGTATTTCTGCTGTGTCTGTTTCTGCCCTTAATTTATATATATATATATATATTTTTTTTTTTTGGAGACA
GAGTCTTGGTCTGTGCGCCAGGGTGAGTGCAGTGGTGTGATCACAGGTGAGTGAACTTTACCTTCTGGCCTGAGCCGT
CCTCTCACCTCAGCCTCCTGAGTAGCTGGAACGCAGACACGACCGCTACACCTGGCTAATTTTAAATTTTCTGGA
15 GACAGGGTCTGTGTGTGTGCCAGGCTGGTCTCAAACCTCTGGACTCAAGGGATCCATCTACCTCGGCTTCCAAAGTG
CTGAATTACAGGCATGAGCCACCATGTCTGGCCTAATTTTCAACACTTTTATATTCTTAGTGTGGGTATGCTCTGTTA
ACAGCATGTAGGTGAATTTCCAATCCAGTCTGACAGTCTGTTGTTAACTGGATAACCTGATTATTTTCTATTTTGTGTC
ACTAGAGACCCGCTGGTGCATCTGATTCTCCACTTGCTGTGTCATGTCCTCGTTCCTTGTCTTCTCACCACCTCTTG
GGTTGCCATGTGCGTTTCTGCGAGTGTGTGATCCTCTCGTTGCTCCTGGTCACTGGGCATTGCTTTTATTCT
20 CTTTGCTTAGTGTACCCCTGATCTTTTATTGTGCTGTGTTGCTTTTGTATTAGACAGTCTCACTCTGTACCCCA
GGCTGGAGTGAATGGACAATCTCGGCTCACTGCAACCTCTGCTCCTCGGTTCAAGCAGTTCTCATTCCTCAACCTCA
TGAGTAGCTGGGATTACAGGCGCCACCACGCTGGCTAATTTTGTATTTTGTAGAGATAGGCTTTCACCATGT
TGGCCAGGCTGGTCTCAAACCTCTGACCTCAAGTGATCTGCCCGCTTGGCCTCCACAGTGCTGGGATTACAGGTGCAA
GCCACCGTGCCCGCATACCTTGATCTTTTAAATGAAGTCTGAAACATTGCTACCTTGTCTGAGCAATAAGACCTT
25 AGTGTATTTAGCTCTGGCCACCCCGAGCTGTGTGCTGTTTCCCTGCTGACTTAGTCTATCTCAGGCATCTTGACA
CCCCACAAGCTAAGCATTAATAATTGTTTTCCGTGTGAGTGTCTGTAGCTTTGCCCGCCCTGCTTTTCTCTC
TTTGTTCCTGCTGTCTTCTGTCTCAGGCCCGCGTCTGGGGTCCCTTCTGTCTTTGCGTGGTCTTCTGTCTTG
TTATTGCTGGTAAACCCAGCTTTACCTGTGCTGGCCTCCATGGCATCTAGCGAGTCCGGGACCTCTGCTTATGATGC
ACAGATGAAGATGTGGAGACTCACGAGGAGGCGGTCTCTTGGCCCGTGAGTGTCTGGAGCACCAGTGGCCAGCGTTC
30 CTTAGCCAGTGAGTGACAGCAACGTCGCTCGGCTGGGTTGAGCTGGAAAACCCAGGCATGTGCGGGTCTGGTGGCT
CCGCGGTGTGAGTTTGAATCGGCAACCTGCGGTGTGGCGCCAGCTCTGACGGTGTGCTGGCGGGGAGTGTCTG
CTTCTCCCTTCTGCTTGGGAACAGGACAAAGGATGAGGCTCCGAGCCGTTGTGCGCCAACAGGAGCATGACGTGAGCC
ATGTGGATAATTTTAAATTTCTAGGCTGGGCGCGGTGGCTCACGCTGTAATCCAGCACTTTGGGAGGCCAAGCGGG
TGGATCACGAGGTGAGGAGGTGAGACCATCTGGCCAACATGATGAAACCCATCTGTACTAAAAACAAAAATTAGC
35 TGGGCGTGGTGGGGGTGCTGTAATCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGTTGGAA
GTTGCAGTGAGCCGACATTGCACCACTGCACTCCAGCCTGGCAACACAGCGAGACTCTGTCTCAAAAAAAAAAAAAA
AAAAAAAAAAATTTAGTAGCCACATTAAAAAGTAAAAAGAAAGGTGAAATTAATGTAATAATAGATTTTACTGAA
GCCCAGCATGTCCACACCTCATATTTAGGGTGTATTGGTGGGAGCATCACTCACAGGACATTTGACATTTTGTAGC
TTTGTCTGCGGATCCCGTGTGTAGTCCCGTGTGGCCATCTCGGCTGGACCTGCTGGGCTTCCCATGGCCATGGCT
40 GTTGTACCAGATGGTGCAGTCCGGATGAGGTGCGCAGGCCCTCAGTGAGCTGGATGTGCAGTGTCCGATGGTGCACG
TCTGGGATGAGGTGCGCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTGAGGGTGAAGTCTCCAG

5 G C C C T C G G T G A G C T G G A G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A G C T G G A T G
T G T G G T G T C T G G A T G G T G C A G G T C A G G G T G A G G T C T C C A G G C C C T C G G T A A G C T G G A G G T A T G G A G T C C G G A T G A T G C A
G G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A G C T G G A T G T G G T G T C T G G A T G G T G C A G G T C T G G G G T G A G G T C A C C
A G G C C C T G C G G T G A G C T G G G T G T G C G G T G T C T G G A T G G T G C A G G T C T G G A G T G A G G T C G C C A G A C G G T G C C A G A C C A T G C
10 G G T G A G C T G G A T A T G C G G T G T C C G G A T G G T G C A G G T C T G G G G T G A G G T T G C C A G G C C C T G C T G T G A G T T G G A T G T G G G G T
G T C C G G A T G C T G C A G G T C C G G T G T G A G G T C A C C A G G C C C T G C T G T G A G C T G G A T G T G T G G T G T C T G G A T G G T G C A G G T C T
G G G G T G A A G G T C G C C A G G C C C T G C T T G T G A G C T G G A T G T G T G G T G T C T G G A T G G T G C A G G T C T G G A T G A G G T C G C C A G
G C C C T C G G T G A G C T G G A T G T G C A G T G T C C A G A T G G T G C A G G T C C G G G T G A G G T C G C C A G A C C C T G C G G T G A G C T G G A T G
T G C G G T G T C T G G A T G G T G C A G G T C T G G A T G A G G T C G C C A G G C C C T C G G T G A G C T G G A T G T A T G G A G T C C G G A T G G T G C C
15 G G T C C G G G T G A G G T C G C C A G A C C C T G C T G T G A G C T G G A T G T G C G G T G T C T G G A T G G T A C A G G T C T G G A T G A G G T C G C C
A G A C C C T G C T G T G A G C T G G A T A T G C G G T G T C C G G A T G G T G C A G G T C A G G G T G A G G T C T C C A G G C C C T C G G T G A G C T G G A
G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A A C T G G A T G T G C G G C G T C T G G A T G G T
G C A G G T C T G G G G T G T G G T C G C C A G G C C C T C G G T G A G C T G G A G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T C G
C C A G G C C C T G C T G T G A G C T G G A T G T G C G G C G T C T G G A T G G T G C A G G T C T G G G G T G T G G T C G C C A G G C C C T C G G T G A G C T G
20 G A G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T T G C C A G G C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G
G T G C A G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G G T G C A G G T C T G G G G T G A G G T
C A C C A G G C C C T G C G G T G A G C T G G T T G T G C G G T G T C C G G T T G C T G C A G G T C C G G G T G A G T T C G C C A G G C C C T C G G T G A G C
T G G A T G T G C G G T G T C C C C G T G T C C G G A T G G T G C A G G T C C A G G G T G A G G T C G C T A G G C C C T T G G T G G G C T G G A T G T G C C G T
G T C C G G A T G G T G C A G G T C T G G G G T G A G G T C G C C A G G C C C T T T G G T G A G C T G G A T G T G C G G T G T C T G C A T G G T G C A G G T C T G
25 G G G T G A G G T C G C C A G G C C C T T G G T G G G C T G G A T G T G T G G T G T C C G G A T G G T G C A G G T C C G G C G T G A G G T C G C C A G G C C C T
G C T G T G A G C T G G A T G T G C G G T G T C T G G A T G G T G C A G G T C C G G G T G A G G T A G C C A A G G C C T T C G G T G A G C T G G A T G T G G G
G T G T C C G G A T G G T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C G G T T A G C T G G A T A T G C G G T G T C C G G A T G G T G C A G G T
C C G G G T G A G G T C A C C A G G C C C T G C G G T T A G C T G G A T G T G C G G T G T C T G G A T G G T G C A G G T C C G G G T G A G G T C G C C A G G
C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G G T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C A G T G A G C T G G A T G
30 T G C T G T A T C C G G A T G G T G C A G G T C T G G C G T G A G G T C G C C A G G C C C T G C G G T T A G C T G G A T A T G C G G T G T C G G A T G G T G C A
G G T C C G G G T G A G G T C A C C A G G C C C T G C G G T T A G C T G G A T G T G C G G T G T C C G G A T G G T G C A G G T C T G G G G T G A G G T C G C C
A G G C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G G T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C G G T G A G C T G G
A T G T G C T G T A T C C G G A T G G T G C A G G T C T G G C G T G A G G T C G C C A G G C C C T G C G G T G A G C T G G A T G T G C A G T G T A C G G A T G G
T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C G G T G G G C T G T A T G T G T G T G T C T G G A T G G T G C A G G T C C G G G T G A G T T
35 C G C C A G G C C C T G C G G T G A G C T G G A T G T G T G G T G T C T G G A T G C T G C A G G T C C G G G T G A G T T C G C C A G G C C C T C G G T G A G C
T G G A T A T G C G G T G T C C C C G T G T C C G A A T G G T G C A G G T C C A G G G T G A G G T C G C C A G G C C C T T G G T G G G C T G G A T G T G C C G T
G T C C G G A T G G T G C A G G T C T G G G G T G A G G T C G C C A G G C C C T T G G T G A G C T G G A T G T G C G G T G T C C G G A T G G T G C A G G T C C G
G G G T G A G G T C A C C A G G C C C T C G G T G A T C T G G A T G T G G C A T G T C C T T C T C G T T T A A G

35 Intron 3 (SEQ ID NO 6)

G T A C T G T A T C C C C A G C C A G G C C T C T G C T T C T C G A A G T C C T G G A A C A C C A G C C C G G C C T C A G C A T G C G C C T G T C T C C A C T
T G C C T G T G C T T C C C T G G C T G T G C A G C T C T G G G C T G G G A G C C A G G G C C C C G T C A C A G G C C T G G T C C A A G T G G A T T C T G T G
C A A G G C T C T G A C T G C C T G G A G C T C A G T T C T C T A C T T G T A A A A T C A G G A G T T T G T G C C A A G T G G T C T C T A G G G T T T G T A
A A G C A G A A G G G A T T T A A A T T A G A T G G A A C A C T A C C A C T A G C C T C C T T G C C T T T C C C T G G G A T G T G G G C T G A T T C T C T C
40 T C T C T T T T T T T T T C T T T T T G A G A T G G A G T C T C A C T C T G T G C C C A G G C T G G A G T G C A G T G G C A T A A T C T T G G C T C A C T

5 GCAACCTCCACCTCCTGGGTTTAAGCGATTACCAGCCTCAGCCTCCTAAGTAGCTGGGATTACAGGCACCTGCCACCAC
GCCTGGCTAATTTTTGTACTTTTAGGAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCATGACCTCAGG
TGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGTTTACAGGCTAAGCCACCGTGCCAGCCCCGATTCTCTTTTAATT
CATGCTGTTCTGTATGAATCTTCAATCTATTGGATTTAGGTCATGAGAGGATAAAATCCACCCACTTGGCGACTCACTG
10 CAGGGAGCACCTGTGCAGGGAGCACCTGGGGATAGGAGAGTTCCACCATGAGCTAACTTCTAGGTGGCTGCATTGAATG
GCTGTGAGATTTGTCTGCAATGTTGCGCTGATGAGAGTGTGAGATTGTGACAGATTCAAGCTGGATTTCATCAGTGAG
GGACGGGAGCGCTGGTCTGGGAGATGCCAGCCTGGCTGAGCCCAGGCCATGGTATTAGCTTCTCCGTGTCCGCCCAGGC
TGACTGTGGAGGGCTTTAGTCAGAAGATCAGGGCTTCCCAGCTCCCTGCACACTCGAGTCCCTGGGGGGCCTTGTGAC
ACCCCATGCCCAAATCAGGATGTCTGCAGAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGGCTGGG
15 GACCCCGACGTGGTGTCTGGGGCCATTTCTTGCATCTGGGGAGGGTCAGGGCTTCCCTGTGGGAACAAGTTAATACAC
AATGCACCTTACTTAGACTTTACACGTATTTAATGGTGTGCGACCCAACATGGTCATTTGACCAGTATTTTGAAAGAAAT
TTAATTGGGGTGACCGAAGGAGCAGACAGACGTGGTGGTCCCAAGATGCTCCTTGTCACTACTGGGACTGTTGTTCTG
CCTGGGGGGCCTTGGAGGCCCTCCTCCCTGGACAGGGTACCGTGCCCTTTTCTACTCTGCTGGGCCTGCGGCCTGCGGT
AGGGCACCAGCTCCGAGCACCCGCGGCCCCAGTGTCCACGGAGTGCCAGGCTGTGAGCCACAGATGCCAGGTCAGGT
20 GTGGCCGCTCCAGCCCCGTGCCCCATGGGTGGTTTTGGGGAAAAGGCCAAGGGCAGAGGTGTGAGGAGTGGTGGG
CTCATGAGAGCTGATTCTGCTCCTTGGCTGAGCTGCCCTGAGCAGCCTCTCCGCCCCCTCTCATCTGAAGGATGTGGCT
CTTTCTACCTGGGGGTCTGCTGCTGGGGCCAGCCTTGGGCTACCCAGTGGCTGTACCAGAGGGACAGGCATCCTGTGTGG
AGGGGCATGGGTTCAGTGGCCCCAGATGCAGCCTGGGACCAGGCTCCCTGGTGTGATGGTGGGACAGTACCCTGGGG
GTTGACCGCCGACTGGGCGTCCCAGGGTTGACTATAGGACCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAG
25 AGGCGTCTGGCTGGCATGGGTGGACGTGGCCCCGGGCATGGCCTTCAGCGTGTGCTGCCGTGGGTGCCCTGAGCCCTCAC
TGAGTCGGTGGGGGCTTGTGGCTTCCCGTGAGCTTCCCCTAGTCTGTTGTCTGGCTGAGCAAGCCTCCTGAGGGGCTCT
CTATTGCAG

Intron 4 (SEQ ID NO 7)

25 GTGGCTGTGCTTTGGTTTAACTTCCTTTTAAACAGAAGTGCCTTTGAGCCCCACATTTGGTATCAGCTTAGATGAAGGG
CCCGAGGAGGGGCCACGGGACACAGCCAGGGCCATGGCAGGCGCCAACCATTTGTGCGCAGTGAGGTGGCCGAGG
TGCCGGTGCCCTCCAGAAAAGCAGCGTGGGGGTGTAGGGGAGCTCCTGGGGCAGGGACAGGCTCTGAGGACCACAAGAAG
CAGCCGGGCCAGGGCCTGGATGCAGCACGGCCCCAGGTCCTGGATCCGTGTCTGTGGTGCAGCCTCCGTGCGCT
TCCGCTTACGGGGCCCGGGGACCAGGCCACGACTGCCAGGAGCCACCGGGCTCTGAGGATCTGGACCTTGCCCCACGG
30 CTCTGCACCCACCCCTGTGGCTGCGGTGGCTGCGGTGACCCCGTCATCTGAGGAGAGTGTGGGGTGGGTGGACAGAG
GTGTGGCATGAGGATCCCGTGTGCAACACACATGCCGCCAGGAACCGTTTCAAACAGGGTCTGAGGAAGCTGGGAGGGG
TTCTAGGTCCCGGTCTGGGTGGCTGGGGACACTGGGGAGGGGCTGCTTCTCCCCTGGGTCCCTATGGTGGGGTGGGCAC
TTGGCCGGATCCACTTTCCTGACTGTCTCCCATGCTGTCCCGCCAG

35 Intron 5 (SEQ ID NO 8)

GTGGGTGCCGGGACCCCCGTGAGCAGCCCTGCTGGACCTTGGGAGTGGCTGCCTGATTGGCACCTCATGTTGGGTGGAG
GAGGTACTCCTGGGTGGGCGCAGGGAGTGACGGTGACCCTGTCACTGTTGAGGACACACCTGGCACCTAGGGTGGAGGC
CTTCAGCCTTTCCTGCAGCACATGGGGCGACTGTGCACCTGACTGCCCGGGCTCCTATTCCAAGGAGGGTCCCACTG
GATTCCAGTTTCCGTGAGAGAAGGAACCGCAACGGCTCAGCCACCAGGCCCCGGTGCCTTGACCCCCAGTCTGAGCCAG
40 GGGTCTCCTGTCTGAGGCTCAGAGAGGGGACACAGCCCCCTGCCCCTTGGGGTCTGGAGTGGTGGGGTCTCAGAGAGAG

5'-Bereich Intron 6 (SEQ ID NO 9)

3'-Bereich Intron 6 (SEQ ID NO 10)

GTGTGGGATTGGTTTTCATGTGTGGGATAGGTGGGGGATCTGTGGGATTGGTTTTCATAGTGGGGTAACACAGAGTTCAGG
GCGAGCTTTCTCCTGTAGTGGGTCTGCAGGTGCTCCAACAGCTTTATTGAGGAGACCATATCTTCCTTTGAACTATGGT
20 CGGGTTTATAGTAAGTCAGGGGTGTGGAGGCCTCCCCTGGGCTCCCTGTTCTGTTCCTTCCACTCTGGGGTCGTGTGGTG
CCTGCTGTGGTGTGTGGCCGGTGGGACAGGGCTTCAGGCCTCCTTGTGTTTCATTGGCCTGGATGTGGCCCTGGCTACGCT
CCGTCTTGGAAATCCCCTGCGAGTTGGAGGCTTCTTTCTTTCTTTTTTCTTTCTTTTTTTTTTTTTTTTGATAACAGA
GTCTCGCTCTTTTTTGCCACAGGCTGGAGTGGTTTGGCGTGATCTTGGCCTACTGCAACCTGTGCTTCCTGAGTTCAAGCA
ATTCTCTTGCCTCAGCCTCCCAAGTAGCTGGAATTATAGGCGCCACCACCATGCTGACTAATTTTTGTAAATTTTAGTAG
25 AGACGAGGTTTCTCCATGTGGCCAGGCTGGTCTCGAACTCTGACCTCAGGTGATCCTCCACCTCGGCCTCCCAAAGT
GCTGGGATGACAGGTGTGAACCGCCGCGCCCGCCGAGACTCGCTTCTGCAGCTTCGTGAGATCTGCAGCGATAGCTG
CCTGCAGCCTTGGTGTGACAACCTCCGTTTTCTTCTCCAGGTCTCGCTAGGGGTCTTTCATTTCATGACTCTCTTCA
CAGAAGAGTTTCACGTGTGCTGATTTCCCGCTGTTTCTGCGTAATTGGTGTCTGCTGTTTATCGATGGCCTCCTTCCA
TTTCTTTAGGCCTTGTATTATGTGTTTTTCCGGCTCCTTGAAGGAAAAGTTTCGATTATGGATGTTTGAACCTTTCTTT
30 TCTAAACAAGCATCTGAAGTTGCCGTTTTCCCTCTAAAGCAGGGATCCCAGGCCCCCTGGCTGTGGAGTGGCACCGGTCT
GGGGCTGTGTAGGAACCCGGCGCACAGCGGGAGGCTAGGTGGGGTGTGGGGAGCCAGCGTTCGCCCTGAGCCCGCCCC
TCTCAGATCAGCAGTGGCATGCGGTGCTCAGAGGCGCACACCCCTACTGAGAACTGTGCGTGAGAGGGGTCTAGATTCT
GTGCTCCTTATGGGAATCTAATGCTGATGATCTGAGGTGGAACCGTTTGCTCCCAAACCATCCCCCTCCCCACTGCTG
TCCTGTGGAAAAATCGTCTTCCACGA AACAGTCCCTGGTACCACAATGGTTGGGGACCCTGTGCTAAAGACCTGCTTCA
35 GCAGCCTCTCGTCAGTGTGATATATTGGCTTTTCTGTGTTGAGTCCAGAATAATTACGGATTTCGTGTATGCTTTCCGC
CGACCTCAGACCCATGGGCTATTGTGGGCGTGTGCTGCTCCTGGGTGGGAAGGGTGCAGGCCCCATGTACCTTCTCT
GTTACTGCCTTCCAGGTGGTTCTCAGGGTTGAATCGTACTCGATGTGGTTTTAGCCACGGCCCTGCCGCCAGCTCCTG
GGGGCTGGGGAACATGCTGAAGCACAGAGTCACCGTGC GCGCTCTTTTGTATGCTCACAAGCTCGAGGCCTCCTGTGTCG
TGTTAGTGTGTGTACGTGCTGCTCACATCCTGTCTTGGGGACGACAGGGGCTTAGCAGGTCCCCTAGTAAATGACAAGC
40 GTCTCTGGGGGAGTCTGCAGAATAGGAGGTGGGGGTGCCGCTCTCTCTCCGCGTCTTACAGACTCTTCTCCTGCTGTGCTG

GTGGCTGCACCTGCATCCCTGCAATCCCTCCAGCACTGGGCTGGAGAGGCCCGGGAGCTCGAGTGCCACTTGTGCCACGT
GACTGTGGATGGCAGTCGGTCACGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTGGTCAAGGGGTCTGATGTGTG
GTGACTGTGGATGGCGGTCTGATGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGG
ATGGCGGTCTGATGTGGTGACTGTGGATGGCGGTCTGATGTGGTGACTGTGGATGGCGGTCTGAT
5 GGGTCTGATGTGGTGACTGTGGATGGCAGTCGGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGAT
TGGTGACTGTGGATGGCAGTCGGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACT
GTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGG
CGGTCTGATGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGTGATCGGTCA
CAGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGTGATCGGTCA
10 GGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACT
TCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACT
GATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACT
GACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGAT
GGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACT
15 TCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACT
TCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACT
GGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGTGATCGGTCAAGGGGTCTGATGTGTGGTGACT
GACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGAT
GGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACT
20 GTCAAGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACT
GGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACT
GTGGTGACTGTGGATGGTGATCGGTCAAGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCTGATGTGTGGTGACT
ACTTTGCGCTCTCGGCCCCCGGCCCCGTTTCCCAACAGAAAGCTTCCAGGCGCTCTCTGGGCTTCATCCCGCATCG
GGCTTGGCCGAGGTCCACAGTCTGATCGGAAGAAACAAGTCCAGCTCTGGCCGGGCGAGGCCACATTTGTGGCTC
25 ATGCCCTCTCCTCTGCCGGCAG

Intron 7 (SEQ ID NO 11)

GTCTGGGCACTGCCCTGCAGGGTTGGGCACGGACTCCAGCAGTGGGTCTCCCTGGGCAATCACTGGGCTCATGACCG
GACAGACTGTTGGCCCTGGGGGGCAGTGGGGGAATGAGCTGTGATGGGGGCATGATGAGCTGTGTGCCTTGGCGAAATC
30 TGAGCTGGGCCATGCCAGGCTGCAGAGCTGCTGCAATCAGGCACCTGCTCACGTTGACTGCGCGGCTCTCTCCAGTT
CCGAGTGCCCTTGTTCATGATTTGCTAAATGTCTTCTGCGCAGTTTGTGATCTTGAGGCCAAGGAAAGGTGTCCCCCT
CCTTTAGGAGGGCAGGCCATGTTTGAAGCGTGTCTGCCCAGCTGGCCCTCAGTGTGGGTCTGAGGCCAAGGAAACG
TGTCCCCCTTCTTAGGAGGACGGGCGGTGTTTGAAGCACGCCCCGCTGAGCGGGCTCTCAGTGTGGGTCTGTCCACGT
GGCCCTGTGGCCCTTTCAGATGTGGTCTGTCCACGTGGCCCTGTGGCTCTTTCAGATGCCTGTTAGCACTTGCTCGGC
35 TCTAGGGGACAGTCTGTCCACCGCATGAGGCTCAGAGACCTCTGGGCGAATTTCTTGGCTCCAGGGTGGGGGTGGAG
GTGGCTGGGCTGTGGGACCCAGACCTGTGCCCCGCGAGCTGGGCGAGCACTCTGGATCACATATGCCATCCGGGCCA
CGGTGGGCTGTGTGGGTGTGAGCCAGCTGGACCCACAGGTGGCCCGAGGAGACGTTCTGTGTACACACTCTGCCTAA
GCCCATGTGTGTCTGAGAGACTCGGCCCGGCCAGCCACAGTGGCCCTGCATTCCAGCCAGCCCGCACTTCATCACA
AACACTGACCCCAAGGGACGGAGGGTCTGGCCACGTGCTCTGCTGTCTCAGACCCACCGGCTCACTCCCATGTG
40 TCTCCCGTCTGCTTTCGAG

Intron 8 (SEQ ID NO 12)

GTGAGTCAGGTGGCCAGGTGCCATTGCCCTGCGGGTGGCTGGGCGGGCTGGCAGGGCTTCTGCTCACCTCTCTCTGCCC
CTTCCCCACTGNCCTTCTGCCCCGGGGCCACCAGAGTCTCCTTTTCTGGCCCCCGCCCCCTCCGGCTCTGGGCTGCAGGC
5 TCCCGAGGCCCCGAAACATGGCTCGGCTTGGCGCAGCCGGAGCGGAGCAGGTGCCACAGAGCCTGGAAATGGCAAGC
GGGGTGTGGAGTTGCTCCTGCGTGGAGGACGAGGGGCGGGGGTGTGTCTGGGTGAGGTGTGCGCCGAGCGTTTGAGCCT
GCAGCTTGTGAGCTCCAAGTTACTACTGAGCTGGACACCCGGCTCTCACACGCTTGTATCTCTCTCTCCGATACAAAA
GGATTTTATCCGATTCTCATTCTGTCCCTGTGCTGTGACCCCCGAGGGCGGGGCTCTTCTCTCTGTGACTAGATTT
CCCATCTGGAAAGTGGGGGTTGACCGTGTAGTTTGTCTCTCGGGGGGCTGTGGTGGCCATGGGGCAGGCGGCCTGG
10 GAGAGCTGCCGTACACAGCCACTGGGTGAGCCACACTCAGGTGGTAGAGCCACAGTGCTGGTGCCACATCACGTCCT
CTGGATTTTAAGTAAAACACACACTCCCGGAGGCATCTGCTGCGACCCCTGTGTGCTCTGGGAGAGTGGTAGCAC
GGAGGAAATTCGTGCACACTCAAGGTCATCAGCAAGGTCATCCGCAGTCAGGTGGAACGTGGAGGCCTCTCTCTGGGATC
GTCTCCAGCGGATAAAGGACTGTGCACAGCTTCGGAAGCTTTTATTTAAAAATAAACTATTAATTATTGCATTATAAGT
AATCACTAATGGTATCAGCAATTATAATTTTATTAAAGTATAATTAGAAATATTAAGTAGTACACGTTCTGGAAAAA
15 CACAAATTGCACATGGCAGCAGAGTGAATTTGGCCGAGGACACGTGTGCACATGTGTGAAGCGGCCCCAGGCCAC
AGAATTGCTGACAAAGTCACCTCCCCAGAGAAGCCACCAGGGCCTCTTCGTGGTGTGAATTTTATTAAGATGGATC
AAGTCACGTACCGTCCACGTGTGGCAGGGCTTTGGGAATGTGAGGTGATGACTGCGTCTCATGCCCTGACAGACAGGA
GGTGACTGTGTCTGTCTGTCCCTAGGACACGGACAGGCCGAAGCTCTAGTCCCCATCGTGGTCCAGTTTGGCCTCTGA
ATAAAACGTCTTCAAAACCTGTTGCCCCAAAACTAAGAACAGAGAGAGTTTCCCATCCCATGTGCTCACAGGGGCGTA
20 TCTGCTTGGTTGACTCGCTGGGCTGGCCGACTCCTAGAGTTGGTGGTGTGCTTCTGTGCAAAAAGTGCAGTCTCTT
GCCCCATCACTGTGATATCTGCACCAGCAAGGAAAGCCTCTTTTCTTTCTTTCTTTTCTTTTGGAGACGGAACGTCA
CTGTGTCTGCTGGGCTTGAGTGCAGTGGCGGATCTCAACTCACTGCAACCTCCGCTCCCGGTTCCAGCATTTCTC
CTGCCCTCAGCCTCCCGAGCAGCTGAGATTACAGGCACCCACCCCTGCGCCTGGCTAATTTTTGTATTTTAGTAGAGAG
GGGTTTTTGCCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTG
25 GGATTACAGGTGTGAGCCATCACGCCAGCCGAAAGCCTCTTTTAAAGGTACCACCTATAGCGCTTCCCGAAAATAAC
AGGTCTTGTTTTGCAGTAGGCTGCAAGCGTCTCTTAGCAACAGGAGTGGCGTCTGTGGGCTCTGGGATGGCTGAGGG
TCGCGTGGCAGCCATGCCTTCTGTGTGCACCTTTAGGTTCCACGGGGCTATTCTGCTCTCACTGTTTGTCTGAAAACGCA
CCCTTGGCATCCTGTTTGGAGAGTTTCTGCTTCTCGTTGGTCACTGCTGAAACTAGGGGCAAGTTGTATCCGTTGGCGC
GCAGCGGCTACATGTAGGGTCATGAGTCTTTACCGTGGACAAATTCCTTGAAAAAAGGAGTCCGGTTAAGCAT
30 TCATTCCGGGTCAAGTGTCTGGTTCTGTGAATAAATCTAAGATTTAAGAAACCTTAATGAAAGAAACCTTGATGATTC
AGAGCAAGGATGTGGTCACACCTGTGGCTGGATCTGTTTACGCCGCCAGTGTCATGGTGAGAGTGGGAGCAGGGATTG
TTTGTTCAGAGGTCTCATCTGGTATGTTTCTGAGGTGTTTGGCGGCTGAATGGTAGACGTGCTGTTGTGTATGAGGT
TCTGTGTCTGTGTGGCTCGGTTTGTAGGTACGCATGTCCAGCACATGCCCTGCCGTCTCTCACCTGTGTCTTCCCGC
CCAG
35

Intron 9 (SEQ ID NO 13)

GTGAGGCTCCTCTTCCCCAGGGGGCTTGGGTGGGGGTTGATTGCTTTTGATGCATTCACTGTTAATATTCCTGGTGC
TCTGGAGACCATGACTGCTCTGTCTTGAGGAACAGACAAGGTTGCAGCCCCCTTCTGGTATGAAGCCGACGGGAGGGG
TTGCACAGCCTGAGGACTGCGGGCTCCACGAGGCTCTGTCCAGCGGCCATGTCCAGAGGCCTCAGGGCTCAGCAGGCGG
40 GAGGGCCGCTGCCCTGCATGATGAGCATGTGAATCAACACCGAGGAAGCACACCAGCTTCTGTACGTCAACCAGGTTT

CGTTAGGGTCCTTGGGGAGATGGGGCTGGTGCAGCCTGAGGCCCCACATCTCCAGCAGGCCCTCGACAGGTGGCCTGGA
CTGGGCGCCTCTTCAGCCCATTCGCCATCCCACTTGCATGGGGTCTACCCCAAGGACGCACACACCTAAATATCGTGCC
AACCTAATGTGGTTCAACTCAGCTGGCTTTTATTGACAGCAGTTACTTTTTTTTTTTAACTTTAAGTTCAGGGTAC
ATGTGCACGACGTGCAGGTTAGTTACATATGTATACATGTGCCATGTTGGTGTGCTGCACCCATTAACTCATCATTACAA
5 TTAGGTATATCTCCTAATGCTATCCCTCCCCACTCCCCCATCCCATGACAGGCCCTGGTGTGTGATGTTCCCCACCTG
TGTCCAAGTGTCTCATGTTCAGTTCACCTGTGAGTGAGAACATGTGGTGTGGTGTCTTCTCCTTGAATAGTTT
GCTCAGAGTGATGGTTTCCAGCTTCGTCCATGTCCCTACAAAGGACATGAACTCATCCTTTTTATGACTGCATAGTATT
CCGTGGTGTATATGTGCCACATTTTCTTAATCCAGTCTATCATCGATGGACATTTGGGTTGGTTGCAAGTCTTTGCTACT
GTGAATAGTGC CGCAATAAACATACGTGTGCATGTGTCTTTATAGCAGCATGATTATAATCCTTTGGGTATATACCCAG
10 TAATGGGATGGCTGGGTCAAATGGTATTTCTAGTCTTAGATCCTTGAGGAATCACCACACTGTCTCCACAATGTTGAA
CTAGTTTACACTCCCAACAGTGTAAGAGTGTCTGGTGCTGGAGAGGATGTGGACAGCAGTTATTTTTATGAAAA
TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACACAGCCATTCTCTCGAAGAC
TCCGGGTTTTCTGTGCATCTTTGAACTCTAGCTCCAATTATAGCATGTACAGTGGATCAAGGTTCTTCTTCATTAA
GGTTCAAGTCTAGATTGAAATAAGTTTATGTAACAGAAAAAAATTTCTTGTACACAACTGTCTCTGGGATTGGA
15 GGAAAGTGTCTCGAGCTGGCGGCACACTGGTCAGCCCTCTGGGACAGGATACCTCTGGCCCATGGTCATGGGCGCTGG
GCTTGGGCTGAGGGTCACACAGTGACCATGCCAGCTTCTGTGGATAGGATCTGGGTCTCGGATCATGCTGAGGACC
ACAGCTGCCATGTGGTAAAGGGCACACGTGGCTCAGAGGGGGCAGGTTCCAGCCCCAGCTTCTTACCGTCTTCAG
TTATTTTCCCTAAGAGTCTGAGAAGTGGGGCCGCGCTGATGGCCTTCTGTCGTCTTCACTGGCACAGAATTGCACAA
GCTGATGGTAAACACTGAGTACTTATAATGAATGAGGAATTGCTGTAGCAGTTAACTGTAGAGAGCTCGTCTGTTGAAA
20 GAAATTTAAGTTTTCAITTAACCGCTTTGGAGAATGTTACTTTATTTATGGCTGTGTAATGTTTGACATTCACTCCC
TCGTAGACAGATACTACGTAAAAAGTGAAGTTAACTTGCTGTGTATTTCCCTATTTTATG

Intron 10 (SEQ ID NO 14)

GTGAGGCCCGTGCGTGTCTGTGGGGACCTCCACAGCCTGTGGGCTTGCAGTTGAGCCCCCGTGTCTGCCCCCTGG
25 CACCGCAGCGTTGTCTCTGCCAAGTCTCTCTCTCTGCCGTGCTGGATCCGCAAGAGCAGAGGCGCTTGGCCGTGCACC
CAGGCCCTGGGGCGCAGGGGCACCTTCGGGAGGGAGTGGGTACCGTGACGGCCCTGGTCTTGACAGACGCACCCAGGTT
ACACACGTGGTGAGTGACGGCGGTGACCTGGCTCCTGTGCTCTTTGGAAAGTCAAGAGTGGCGGCTCCTGGGGCCCCAG
TGAGACCCCCAGGAGCTGTGACAGGGCCTGCAGGGCCGAGGCGGCAGCCTCCTCCCAGGGTGACCTGAGCCTGCGGA
GAGCAGGAGCTGTGAGTGAGCTGGCCACAGCGTTGCTGCGGTACGTTCTGCTGCGGGTGTGTTGGGATCGGTGGG
30 AGAATTTGGATTGTGAGTGCTGTCTTGAACACGGAGATGGCTAGGAGTGGGTTTCAGAGTTGATTTTGTGAAT
CAAACTAAAATCAGGCACAGGGGACCTGGCCTCAGCACAGGGGATTGTCCAATGTGGTCCCCCTCAAGGGCGCCCCACAG
AGCCGGTGGGCTGTGTTTAAAGTGCATTTGACAGGGACGAGAAACCTTGAAAGCTGTAAGGGGAACCTCAGAAAAATG
TGGCCGCCAGGGGTGGTTTCAGGTGCTTTGCTGGGCTGTGTTGTGAAAACCCATTGGAACCCGCCCTCAAGTCCACCC
TCCAGGTCCACCTCCAGGCGCGCCCTGGGCTGGGGGTATGCCCTGGCGTTCCTTGTGCGCAGCCCGAGCACAGCAGGC
35 TGTGCACATTTAAATCCACTAAGATTCACTCGGGGGAGCCAGGTCCCAAGCAACTGAGGGCTCAGGAGTCTGAGGCT
GCTGAGGGGACAGAGCAGCGGGGAACGTGCTTCTGTGTGGCAAGTTCCTGAGGGTGTGGCCAGGGAGGTGCTCAG
GTGTATGTTGGGTTCCACCGGGGGCAGAACTCTGTCTCTGATGAGTCGGCAGCCATGTAACAGGAAGGGGTGGCCACAG
GGAGCTGGGAATGCACAGGGGAGCTGCGCAGCTGGCCGAGGTCCAGGGCCAGGCCACAGGAAGGGCAGGGGGACGCCC
GGGGCCACAGCAGAGGCCGAGGAAGGGAAGGGATGCCAGGCCAGAGCAGAGGCTACCGGCCACAGGGGGGCTCCCTG
40 AGCTGGGTGAGCGAGGCTCATGACTCGGCGAGGGAACCTCCTTGACGTGAAGCTGACGACTGGTGTGCCCAGCTCACAG

CCCAGCCAGGTCCCGCGCCTGAGCAGGAACTCAGAACCCTCCCCTTTGTCTAAAGCACAGCAGATGCCTTCAGGGCATCT
AGGAGAAAACAGGCAAAGTCGTTGAGAAACGTCCTTAAAGAAGGTGGGATGGTGGCAATTTCTGTCCAGATTTTAGTCT
GCCCCGGACCACAGATGAGTCTATAACGGGATTGTGGTGTGGCATGGGGACACATGAGATGGACCATCACAGAGGCCAC
TGGGGCTGCACCTCCCATCTGAGTCTTGGCTGTCCCGGGTCCAGGCCAGGTTCTTGATGCTCACCTACCTGTCTGCCCC
5 GGGAGACAGGGAAAGCACCCCGAAGTCTGGAGCAGGGCTGGGTCCAGGCTCCTCAGAGCTCCTGCCAGGCCACGACCCCT
GCTCCAAATCACCACTTCTCTGGGGTTTTCCAAAGCATTTAAACAAGGGTGTGAGTTACCTCCTGGGTGACGGCCCCGCA
TCCTGGGGCTGACATTGCCCTCTGCCTTAG

Intron 11 (SEQ ID NO 15)

10 GTGAGCGCACCTGGCCGGAAGTGGAGCCTGTGCCCGGTGGGGCAGGTGCTGCTGCAGGGCCGTTGCGTCCACCTCTGCT
TCCGTGTGGGGCAGGCGACTGCCAATCCCAAAGGTGAGAGGCCACAGGGTGCCCTCGTCCCATCTGGGGCTGAGCAGA
AATGCATCTTTCTGTGGGAGTGAGGGTGCTCACAACGGGAGCAGTTTTCTGTGCTATTTTGGTAAAAGGAAATGGTGAC
CAGACCTGGGTGCACTGAGGTGTCTTCAGAAAGCAGTCTGGATCCGAACCCAGACGCCCGGGCCCTGCTGGGCGTGAGT
CTCTCAAACCCGAACACAGGGGCCCTGCTGGGCATGAGTCCCTCTGAACCCGAGACCTGGGGCCCTGCTGGGCGTGAGT
15 CTCTCCGAACCCAGAGACTTCAGGGCCCTTTTGGGCGTGAGTCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC
AGTCTACAGGATGCCATGAGTTCATGATCACGTGTACCCATCAGGGGACAGGGCCATGGTGTGGGGGGGTCTCTACAA
AATTCTGGGGTCTGTGTTCCCCAGAGCCGAGAGCTCAAGGCCCGTCTCAGGCTCAGACACAAATGAATTGAAGATGGA
CACAGATGCAGAAATCTGTGCTGTTCTTTTATGAATAAAAAGTATCAACATTCCAGGCAGGGCAAGGTGGCTCACACCT
ATAATCCAGCACTTTGGGAGGCCGAGGTGGGTGGATCACTTGAGGCCAGGAGTTTGAGGCCAACCTAACCAACATAGTG
20 AAATTCATTTCTACTTAAAAAATACAAAAATTAGCCTGGCCTGGTGGCACACGCCCTGTAGTCCCCGTATGCGGGAGGC
TGAGGCAGGAGAATCATTTGAACCCAGGAGGCAGAGGTGTCAGTGAGCCGAGATCACACCACTGCACTCCAGCCCTGGGCA
ACAGAGTGAGACTTCATCTTAAAAAAGTATCAGCATTCCAAAACCATAGTGAGACAGGTGTTTTTTTATTC
TGTCTTCGATAATATTACTGGTGTGTGCTAGAGGCCGGAACCTGGGGGTGCTTCTCTGAAAGGCACACCTTCATGG
GAAGAGAAATAAGTGGTGAATGGTTGTTAAACAGAGGTTTAAACTGGGGTCTGTGTTCTGAGTTAACAGTCCAGATC
25 TGGACTTTGCCTCTTCCAGAATGCTCCTGGGGTTTGCTTCATGGGGAGCAGCAGGTGGACACCTCGTGATGGGG
GAGCAGCAGGTGCAGACGCCCTCATGATGGGGAGTGGCAGGTGCAGACACCTTGTGATGGTGCCAGCATGTCCCTG
TTGAGCTCCCTCCCCACAAGGATGCCGGTCTCCTGTGCTCCCCACAGTCCCTGCTTCCCTCTCACAGCCTTACCTGGTC
CTGGCCTCCACTGGCTTTGTCTGCATGATTTCCACATTTCTGGGCTCCAGCACCTCTTCGCCTCTCCAGGCACCTCT
GCAGTGCTGGCCATACCACTGAGTGTGAAGTGTCCACTGCTTATTTTGTCTCCCATGAAATGTATTTTTTAGGACAGGC
30 ACCCTGTTCCAGCCTCTGGCACAGCATCAGTGAATGTTATTGAAGGACAAAGGACAGACAAACAAATCAGGAAAATGG
GTTCTCTCTAAACACATGCAAAGCCACAGAGGCTAGTGAGGATGGGTGGGCATCAGGTGATCAGATGTGGGTCCAATG
CCAGAAATATTCTGTGCTCCCAAAGGCCACTTGGTCAGAGTGTGTGCTTGAGAGGTGGCTCTAAAGCTCAGCAGTGGAG
GCAGTGGTTCGCCATACTCAGGGTGAAGTACATCCTCTGTGTCTGAAGTATACAGCAGAGGCTTGAAGGGCATCTGGGA
GAAGAAAACAGGCAAAATGATTAAGAAAAGTGAAGGAAAAGTGGTAAGATGGGAATTTCTGTCCAGATTTTAGTCT
35 TCCCAAACCAAGCTCAGATGGTAGAATGTGGTCAGAACTGATGGACAGAAACAAAGAACAAACGGAAGCCCTATCTCT
CAGAAACGTGTGTTAATGTGGTATGTGGCACAGCTGATGGAAAAGAGAGTGTGTGTGTAATTTTTTTTCTGAGAAAAC
GACTGGAAGCAAATAAGTTGTGCTTTACAGCATATACCAGAGCAGATTCTAGGTAGAAGAGGAGACATGCAACAAC
ACCAGCAACAGAAATAAAACAAAGACTCAAAGGGAAGGGAGGTGAACGTTCCCTGGTTTGGTGTGGGGAAGGACACAC
AGGGAGGCGGATGAAACCAAGTGAAGCAACGGGCATTGCTTTCACTGCAGAGAAACTCAGCTTGCTGAGCCACAGTGAAA
40 ATGGCCATTCCCTGGAGCGTTGTGCACGTGATTTATTTAAGGCGCCCTGTGAGGTCTGTGACATTATCCTCTCACTTT

5 GTTCTCCTAACCACCTGAGAGGTAGAGGAGGAAAGGCTCCAGGGGAGCAGCCGCCCTTGGTCACCCAGCTGGCAAAGGSC
ATGCATGATTGCAGCCTGGCCTCCTGCTCCGGGGCCCTTGCTCTGCCCCGAGGACCCACACAAGTCAGACCCATAGGCTC
AGGGTGAGCCGGAGCCCAAGGTCGTGTTGGGGATGGCTGTGAAAGAAGAAATGGACGTCTGATGCACACTTGGGAAGGTC
CTACCAGCAGCGTCAAAGAAATGCATGTGAAACTGACAGCGAGACCCATCCCTCAAAGAAACGCACGTGAAACTGATGGC
10 GAGACCTGTCCCCATCCCTCATGCTGGCTCCTTTCTGGGCTTGCCAAGAGCCAGCATCAGGTTGAGGCAAGCTGGAAAG
ACTTTTCTGGAAAGCAGCTTGTGTCATGGAAGTCTCACAATGTCTGTCTTCCAGTAATTCACCTTCTGAAGTGA
CCAGACATTATCAGGGTCTTATTTACCATTTCCAGTGTTCAGGCAGGGGGACTTGCCACAGCAAGTCACGAACCTGCC
CAAATACAGGGCTAAGGAGATATTATGCATCACAACCTTGCTCTGCCATTAAACATTTTCAAAGAATTTTGAAGAAT
GTTTAAATGGCACAAAACGTTTATTTCAATGTAGCAGTGTCAAAGCTGGATGTAAAAGAACACACCCAGGAGCCTGCCG
15 TGAATGTATGTGTGTTTCTTTGGACATGGACATACATGGGCAGTGAGTGGTGGTGAGGCCCTGGAGGACATCGGTGG
GATGCCTCCATCCTGCCCTCTGGAGACACCATGTGTGCCAGTGCCTCACTGGAGCCCTGTTAGCTGGTGCCACCTG
GCTCTTCCATCCCTGAGATTCAAACACAGTGAGATCCCCACGCCCACTCAGTGTCTCCACAAAAACCTGAGTCAC
ACCTGTGTTCACTCGAGGGACGCCCGGGAGCCAGGGCTCCACAGTTTATTATGTGTTTTGGCTGAGTTATGTGCAGATC
TCATCAGGGCAGATGATGAGTGCACAAACACGCCGTGCGAGGTTTGGATACACTCAACATCACTAGCCAGGTCTGGTG
20 GAGTTTGGTCATGCAGAGTCTGGATGGCATGTAGCATTGGAGTCCATGGAGTGAGCACCCAGCCCCCTGGGCTGCAGC
GCATGCCCCAGGCAGGACAAGGAAGCGGGAGGAAGGCAGGAGGCTCTTTGGAGCAAGCTTTGAGGAGGGGGCTGGGTGT
GGGGCAGGCACCTGTGTCTGACATTCCCCCTGTGTCTCAG

Intron 12 (SEQ ID NO 16)

20 GTGAGCAGGCTGATGGTCAGCACAGAGTTCAGAGTTCAGGAGGTGTGTGCGCAAGTATGTGTGTGTGTGTGCGCGCT
GCCTGCAAGGCTGATGGTGACTGGCTGCACGTAAGAGTGACATGTACGCATATACAGTGAGCACATACATGTGTGCAT
GTGTGTACATGAAGGCATGGCAGTGTGTGCACAGGTGTGCAAGGGCACAGTGTGTGCACATGCCAATGCACACCTGACA
TGATGTGTGTTCTGTGCACAGTCGTGTGGGCATTACGTGAGGTGCATGCGTGTGGGTGTGCAGTGTGAGTAGCATGTGT
GCACATAACATGTATTGAGGGTCTCTGTTTACCCCCGTAGGTCTCAGCACAGTGGCACTCCTTACAGGATGAGAC
25 GGGGTCCCAGGCCTTGGTGGGCTGAGGCTCTGAAGCTGCAGCCCTGAGGGCATTGTCCCCTCTGGGCATCCGCGTCCACT
CCCTCTCCTGTGGGCTTCTGTGTCCACTCCCCCTCTCCTGTGGGCATTACATCCACTCCACTCCCTCTCCTGTGGGC
ATCCGCGTCCACTCCCCCTCTCTGTGGGCATCTGCGTCCACCTCCCCCTCTCTGTGGGCATTGCGTCCACTCCCTCTCCT
GGTTCCTTCTGTCTTGGCCGAGCCTCGGGGGCAGGCAGATGACACAGAGTCTTGAATCGCCAGGGTGGTTTCGACAGTG
30 CCGGTGAGGGCCAGGCCGATTTCACTGGGAAGAGGATAGTTTCTTGTCAAAATGTTCTCTTCTTGTTCATCTGA
ATGGATGATAAAGCAAAAGTAAAACTTAAATCCCAGAGAGGTTTCTACCGTTTCTCACTCTTCTTGGCGACTCTAG

Intron 13 (SEQ ID NO 17)

35 GTGAGCCGCCACCAAGGGGTGCAGGCCAGCCTCCAGGGACCCTCCGCGCTCTGCTCACCTCTGACCCGGGGCTTCACCT
TGGAACCTCTGGGTTTTAGGGGCAAGGAATGTCTTACGTTTTTCAAGTGGTGTCTGCTGCTGTGCACAGTTCTGTTCCGCGT
GCTCTGTGCAAAGCACCTGTTCTCCATCTCTGGGTAGTGGTAGGAGCCGGTGTGGCCCCAGGTGTCCCCACTGTGCCTGT
GCACTGGCCGTGGGACGTATGGAGGCCATCCAGGGCAGCAGGGCATGGGGTAAAGAGATGTTTATGGGGAGTCTTAG
CAGAGGAGGCTGGGAAGGTGTCTGAACAGTAGATGGGAGATCAGATGCCCGAGGATTGGGGTCTCAGCAAAGAGGGCC
GAGGTGGGTGCAGGTGAGGGTCGTGGCCCCACCCCGGGAAGGTGCAGCAGAGCTGTGGCTCCCCACACAGCCCGCCA
GCACCTGTGCTCTGGGCATGGCTGTGCTCCTGGAACGTTCCCTGTCTGGCTGGTCAAGGGGTGCCCTGCCAAGAATCG
40 ACAACTTATCACAGAGGGAAGGGCAATCTGTGGAGGCCACAGGGCCAGCTTCTGCCTGGAGTCAGGGCAGGTGGTGGC

ACAAGCCTCGGGGCTGTACCAAGGGCAGTCGGGCACCACAGGCCCGGGCCTCCACCTCAACAGGCCTCCCGAGCCACTG
GGAGCTGAATGCCAGGAGGCCGAAGCCCTCGCCCATGAGGGCTGAGAAGGAGTGTGAGCATTGTGTATCCAGGGCCG
AGGCTGCGCGAATTACCGTGACACTTGATGTGAAATGAGGTCGTCGTCTATCGTGAAACCCAGCAAGGGCTCACGGGA
GAGTTTTCCATTACAAGGTCGTACCATGAAAATGGTTTTTAACCCGAGTGCTTGCGCCTTCATGCTCTGGCAGGGAGGGC
5 AGAGCCACAGCTGCATGTTACCGCCTTTGCACCAGCTCCAGAGGCTTGGGACCAGGCTGTCTCAGTTCAGGGTGCGTCC
GGCTCAGACCGCCCTCCTCTCTGCTCTCTCTCTGCTCAAATCTTCCCTCGTTTGCATCTCCCTGACGCGTGCTGGG
CCCTCGTGCAAGCTGCTTGACTCCTTTCCGGAACCCCTTGGGGTGTGCTGGATACAGGTGCCACTGAGGACTGGAGGTGT
CTGACACTGTGGTTGACCCAGGGTCCAGCTGGCGTGCTTGGGGCCTCCTTGGGCCATGATGAGGTGAGGAGGATTTTCC
CAGGTGAAAACCTCTGGGAACTCCAGGGCCATGTGACCTGCCACCTGCTCCTCCCATATTAGCTCAGTCTTGTCTCTC
10 ATTTCCACCAGGGTCTCTAGCTCCGAGGAGCTCCCGTAGAGGGCTGGGCTCAGGGCAGGGCGGCTGAGTTTCCCCAC
CCATGTGGGGACCCCTTGGGTAGTCGCTTGATTGGGTAGCCCTGAGGAGGCCGAGATGCGATGGGCCACGGGCCGTTTCCA
AACACAGAGTCAGGCACGTGGAAGGCCCAGGAATCCCTTCCCTCGAGGCAGGAGTGGGAGAACGGAGAGCTGGGCCCCG
ATTTACGGCAGCCAGGCTGCAGTGGGCGAGGCTGTGGTGGTCCACGTGGCGCTGGGGGCGGGGTCTGATTCAAATCCGC
TGGGGCTCGGCCTTCTGGCCCGTGCTGGCCGCGCTCCACACGGGCTTGGGGTGGACGCCCGACCTCTAGCAGGTGGC
15 TATTTCTCCCTTTGGAAGAGAGCCCTCACCCATGTAGGTGTTTCCCTCCTGGGTGAGGAGCGTGGCCGTGTGGCAACC
CCGGGACCTTAGGCTTATTTATTTGTTTAAAAACATTCTGGGCCTGGCTTCCGTGTGTGCTAAATGGGAAAAGACATCC
CACCTCAGCAGAGTTACTGAGAGGCTGAAACCGGGTGCTGGCTTGAAGGTGTGATCTCAGGTCAATCCAGAAGTGGCT
CAGGAAGTCAGTGAGACCAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGTGGGT
GAGGCCAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGAGGGTCAGACCAGGTAC
20 ACGGGGGCTCTGATCACACGCACATATGAGCACATGTGCACATGTGCTGTTTCATGGTAGCCAGTCTGTGCACACCTGC
CCCAAAGTCCCAGGAAGCTGAGAGGCCAAAGATGGAGGCTGACAGGGCTGGCGGGTGGCTCACCTGTAGTCCAGCA
CTTTGGGAGGCCGAGGCGAGAGGATCCCTTGAGCCAGGAGTTTAAGACCAGCCTGAGCAACATAGTAGAACCCCATCTC
TATGAAAAATAAAAACAAAATAGCTGAACATGGTGGTGTGCGCCTGTAGTTCCAATACTTGGGAGGCTGAAGTGGGAG
GATCACTTGAGCCCAGGAGGTGGAAGCTGCAGTGAGCTGAGATTGCACCACTGTACTGCAGCCTGGGTGACAGAGTGAGA
25 GCCCATCTCAACAACAACAAGAAGACTGACAAATGCAGTTTCTTGGAAAGAAACATTTAGTAGGAACTTAACCTACACA
CAGAAGCCAAGTCGGTGTCTCGGTGTGCTGAGATGAGATGATGGGTCTCACACCATACCCAGACCCAGGGTTTATG
CACACAGGGGCGGGTGGCTCAGAAGGGATGCGCAGGACGTTGATATACGATGACATCAAGGTTGTCTGACGAAGGGCAG
GATTCATGATAAGTACCTGCTGGTACACAAGGAACAATGGATAAACTGGAAACCTTAGAGGCCTTCCCGGAACAGGGGCT
AATCAGAAGCCAGCATGGGGGGCTGGCATCCAGGATGGAGCTGCTTCAAGCCTCCACATGCGTGTTCATACAGATGGTGCA
30 CAGAAACGCAGGTACCTGTGCACACACAGACACGCAGCTACTCGCACACACAAGCACACACAGACATGCATGCATGC
ATCCGTGTGTGTGCACCTGTGCCATGAGGAAACCATGCATGTGCATTATGCACGCACACAGGCACCCGGTGGGCCAT
GCCACACCCACGAGCACCGTCTGATTAGGAGGCCTTTCCTCTGACGCTGTCCGCCATCCTCTCAG

Intron 14 (SEQ ID NO 18)

GTATGTGCAGGTGCCTGGCCTCAGTGGCAGCAGTGCCCTGCCTGCTGGTGTAGTGTGTCAGGAGACTGAGTGAATCTGGG
35 CTTAGGAAGTTCTTACCCCTTTTCGCATCAGGAAGTGGTTTAACCCAACCACTGTGAGGCTCGTCTGCCGCCCTCTCGT
GGGGTGAGCAGAGCACCTGATGGAAGGGACAGGAGCTGTCTGGGAGCTGCCATCCTTCCCACCTTGCTCTGCCTGGGGAA
GCGCTGGGGGGCCTGGTCTCTCTGTTTGCCCCATGGTGGGATTTGGGGGGCCTGGCCTCTCTGTTTGCCCTGTGGTGG
GATTGGGCTGTCTCCGCTCCATGGCACTTAGGGCCCTTGTGCAAACCCAGGCCAAGGGCTTAGGAGGAGGCCAGGCCACG
40 GCTACCCCAACCCTCTCAGGAGCAGAGGCCGCGTATCACACGACAGAGCCCCGCGCGCTCTCTGCTTCCAGTCAACG

TCCTCTGCCCCCTGGACACTTTGTCCAGCATCAGGGAGGTTTCTGATCCGTCTGAAATTCAGGCCATGTCGAACCTGCGGT
CCTGAGCTTAACAGCTTCTACTTTCTGTTCTTCTGTGTTGTGGAAATTCACCTGGAGAAGCCGAAGAAAACATTTCTG
TCGTGACTCCTGCGGTGCTTGGGTGCGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCGGGTGTGGGCAGCTTTCCG
GTGTCTCCTGGGAGGGGAGCTGGGCTGGGCTGTGACTCCTCAGCCTCTGTTTTCCCCCAG

5

Intron 15 (WEQ ID NO 19)

GCAAGTGTGGGTGGAGGCCAGTGCAGGCCCCACCTGCCCAGGGGTCACTCTGAACGCCCTGTGTGGGGCGAGCAGCCTC
AGATGTCTGCTGAAGTGCAGACGCCCCCGGCTGACCTGGGGGCTGGAGCCACGCTGGCAGCCCTATGTGATTAAACG
CTGGTGTCCCCAGGCCAGGAGCCTGGCAGGGTCCCCAACTTCTTGAACCCCTGCTTCCCATCTCAGGGCGATGGCTCC
CCACGCTTGGGAGCCTTCTGACCCCTGACCTGTGCTCTCTCAGCCTCTTCCCTGGCTGCTGCCCTGAGCTCCTGGGGT
CCTGAGCAAGTTCTCTCCCGCCCCGCGCTCCAGCGTCACTGGGCTGCTGTCTGCTCGCCCCGTTGGAGGGGTGTCTG
TCCCTTCACTGAGGTTCCACAGCCAGGGCCAGAGGTGCAGGCCCTGCTGCCCCGCCACCCACAGCTCCTAGGAGGG
TTGGAGGATGCCACCTCTGGCTCTTCTGGAACGGAGTCTGATTTTGGCCCCGAG

10

3'-untranskribierter Bereich (SEQ ID NO 20)

ATCTCATGTTGAATCCTAATGTGCACTGCATAGACACCAGTATGCAATTACAGAAGCCTGTGAGTGAACGGGGTGGT
GGTCAGTGCAGGCCCCATGGCCTGGCTGTGCATTACGGAAGTCTATGAGTGAATGGGGTGTGGTGTGAGTGCAGGCCCCATG
GCCTGGCTGGGCTGGGAGGTTTCTGATGTGTGAGGAGGAGGGGAAGGAGGGTAGGGGATAGACAGTGGGAGCCCCCA
CCCTGGAAGACATAACAGTAAGTCCAGGCCGAAGGGCAGCAGGGATGCTGGGGGCCAGCTTGGGCGGCGGGATGATG
GAGGGCCTGGCCAGGGTGGCAGGGATGATGGGGCCCCAGCTGGGGTGGCAGGGGTGATGGGGGGGGCTGGTCTGGGTGG
CGGGGAAGATGGGAAGCCTGGCTGGGCCCCCTCTCCCTGCCTCCACCTGCAGCCGTGGATCCGGATGTGCTTCCCT
GGTGACATCCTCTGGGCATCAGCTTTCATGGAGTGGGGGGCAGGGGCATGACACCATCCTGTATAAAATCCAGGATT
CCTCCTCCTGAACGCCCAACTCAGGTTGAAAGTCAATTCCGCTCTGGCCATTCTCTTAAGAGTAGACCAGGATTCTG
ATCTCTGAAGGTGGGTAGGGTGGGGCAGTGGAGGGTGTGGACACAGGAGGCTTCAGGGTGGGGCTGGTGTGCTCTCTC
ATCCTCTTATCATCTCCAGTCTCATCTCTCATCTCTTATCATCTCCAGTCTCATCTGTCTCTCTTATCTCCAGT
CTCATCTGTCTCTTACCATCTCCAGTCTCATCTCTTATCCTCTTATCTCCTAGTCTCATCCAGACTTACCTCCCA
GGGCGGGTGGCAGGCTCGCAGTGGAGCTGGACATACGTCTTCTCAGGCAGAAAGAACTGGAAGGATTGCAGAGAACAG
GAGGGGCGGCTCAGAGGGACGCAGTCTTGGGGTGAAGAAACAGCCCTCCTCAGAAGTTGGCTTGGGCCACACGAAACCG
AGGGCCCTGCGTGAGTGGCTCCAGAGCCTTCAGCAGGTCCCTGGTGGGGCCTTATGGTATGGCCGGGTCTACTGAGTG
CACCTTGGACAGGGCTTCTGGTTGAGTGCAGCCCGACGTGCCTGGTGTGGGGTGGGGGCTTATGGCCACTGGATATG
GCGTCATTTATTGCTGTGCTTCAGAGAATGTCTGAGTGACCGAGCCTAATGTGTATGGTGGGCCCAAGTCCACAGACTG
TGTCGTAAATGCACTCTGGTGCCTGGAGCCCCGTATAGGAGCTGTGAGGAAGGAGGGGCTCTTGGCAGCCGGCTGGGG
GCGCCTTTGCCCTGCAAACTGGAAGGGAGCGGCCCGGGCGCGTGGGCGGACGACCTCAAGTGAGAGGTGGACAGAAC
AGGGCGGGGACTTCCAGGAGCAGAGGCCGCTGCTCAGGCACACCTGGGTTTGAATCACAGACCAACaGGTCAGGCCATT
GTTGAGCTATCCATCTTCTACAAAGCTCCAGATTCTGTCTTCTCCGGGTGTTTTTTGTTGAAATTTTACTCAGGATTACT
TATATTTTTTGCTAAAGTATTAGACCCTTAAAAAAGGTATTTGCTTTGATATGGCTTAACCTACTAAGCACCTACTTTAT
TTGTCTGTTTTTATTATTATTATTATTATTATTAGAGATGGTGTCTACTCTGTCAACCCAGGTTGTAGTGCAGTGGCAC
AGTCATGGCTCGCTGTAGCCGCAAAACCCAGGCTCAAGTGATCCTCCGGCTCAGCTTCCAGAGTGTGGGATTACAG
GTGTGAGCCACTGCCCTTGCTGGCACTTTTAAAAACCACTATGTAAGGTCAGGTCCAGTGGCTTCCACACCTGTCTATCC
CAGTAGTTTGGGAAGCCGAGGCAGAAAGGATTGTCTGAGGCCAGGAGTTTGGAGCCAGCATGGGTAACATAGGAGACCCC

15

20

25

30

35

40

ATCTCTACAAAAATGCAAAAAGTTATCCGGGCGTGGGGTCCAGCATCTGTAGTCCCAGCTGCTCGGGAGGCTGAGTGGG
AGGATCGCTTGAGCCCGGAGGTCATGGCTGCAGTGAGCTGTGATTGTACCATCGCACTCCAGCCTGGGCAACAGAGTGA
GACCCTGTCTCAAAAAAAAAAAAAAAAAAAGAAGGAGAAGGAGAAGAGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAG
GAAGAAGGAAGAAAGAAGGAGAAGGAGGCGCTGCTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAAATGAAAAAACA
5 AAGTTTTAAAGGGAAGAAAAACCCAGCTCTTTGGACTTCCTTAGGCCTGAACTTCATCTCAAGCAGCTTCCTTCCACA
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGAGGAGAAGCAGGCAAGGGTGGAGGCTGTGGGTGACACCA
GCCAGGACCCCTGAAGGGAGTGGTTGTTTTCTGCCTCAGCCCCACGCTCCTGCCGGTCTGCACCTGCTGTAACCGTC
GATGTTGGTGCCAGGTGCCACCTGGGAAGGATGCTGTGCAGGGGGCTTGCCAACTTTGGTGGGTTTCAGAAGCCCCAG
GCACTTGTTGGCAGGCACAATTACAGCCCTCCCCAAAGATGCCACGTCCTTCTCCTGGAACCTGTGAATGTGTACCCG
10 CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTCATCCTGGATTATCTGG
TGGGCTGATATGGCCACAAGGGTCCCTAGAAGTGAGAGAGGAGGCAGGGGAGAGTCAGAGAGGGGACGTGAGAAGGAC
CACTGGCCACTGCTGGCTTTGAGATGGAGGAGGGGTCCCCAGCCAAGGAATGGGGGCAGCCGCTCCATGCTGGAAAAGC
AAGCAATCCTCCCCGGTCTGAGGGCACACGGCCCTGCCACGCTCGATTTCAGGCCAGTGGGACCTGTTTCAGCTTTC
CGGCCTCCAGAGCTGTAAGATGATGCGTTTGTGTTAGCCACTAAGCTGCAGTGATTGTCACAGCAGCAAATGGAATAG
15 CAGTACAGGGAAATGAATACAGGGACAGTTCTCAGAGTGAATCTCAGCCACCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

10

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

15

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorenght. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des offenen Leserähmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

25

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalisierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

30

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, läßt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion
5 von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplitt wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4
10 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer internen 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

15 Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für
20 Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

25 Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind.
30 Darüberhinaus wurde im Intron 1 eine Sp1-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA⁺-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur
10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv
markierter Primer (5'-GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8x10⁵
cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl
inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL),
2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5 µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl
15 ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser
erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die
Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach
Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden
2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150 µl
20 Phenol/Chloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-
Acetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min
bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol
gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stoplösung gelöst. Nach 5
min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel
25 aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom
ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in
Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA₋₁TTGT)
30 repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA_nNa/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

5 In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losten Promotoren (Smale, 1997) einzuordnen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei 10 anderen TATA-losten Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

Beispiel 7

15 Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.

20 Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem 25 Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer 30 zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer

angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer C5S: 5'-CTGCGTCCTGCTGCGCACGTGGGAAGC-3', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCCTTCCTTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCTCCGCGC-3', 5'-flankierende Region -191 - -171; Primer RP-3A: 5'-CTAGGCCGATTCGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCACGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprinten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, C5S-C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

Beispiel 8

In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschieden lange hTC-Promotorsequenzabschnitte generiert und 5' vor das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCGCTGCCTGAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTTCC-3', Position -4093 bis -4070) wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaares PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGGATTTGCAG-3', Position -3120 bis -3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination PK-3A und PK-5D (5'-

GGAGATCTGATCTTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifiziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmentes (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIA quick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 α -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

Beispiel 10

Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

- 5 CHO-K1 und HEK 293 Zellen wurden in wassergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO₂ kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH₂PO₄, pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit
10 einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszu-plattieren.

- Für die Transfektion wurden pro Well jeweils 2×10^5 HEK 293 Zellen in einer 24-well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium
15 entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serum-freies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium
20 hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H₃PO₄; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl
25 Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO₃)₄ Mg(OH)₂·5H₂O; 2,67 mM MgSO₄; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht gemessen.

- 30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysats und β-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl₂;

50 mM β -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100 μ l 1 M Na_2CO_3 gestoppt und die Absorption bei 420 nm bestimmt.

- 5 Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotorsequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK
10 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im
15 Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

Literaturverzeichnis

- 5 Allsopp, R. C., Vazire, H., Pattersson, C., Goldstein, S., Younglai, E.V., Fitcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.
- 10 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). Current protocols in molecular biology. Greene Publishing Associates and Wiley-Intersciences, New York.
- Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. Nature Genetics 12, 200-204.
- 15 Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.
- Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.
- 20 Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. Science 269, 1236-1241.
- 25 Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthase promoter. Mol. Cell. Biol 14:4894.
- 30 Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.
- 35 Harley, C.B., Fitcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.

- Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346, 866-868.
- 5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* 155, 3711-3715.
- 10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011-2015.
- Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.
- 15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276: 561-567.
- Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633-643.
- 20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234-282.
- 25 Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc. Natl. Acad. Sci.* 86, 7049-7053.
- Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* 41, 181-190.
- 30 Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* 75, 729-739.
- Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* 15, 7155-7174.
- 35 Smale, S.T. and Baltimore, D. (1989). The „initiator“ as a transcription control element. *Cell* 57:103-113.

Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.

5 Shay, J. W. (1997). Telomerae and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.

Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Landsdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with
10 age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.

Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995). Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Journal* 311, 549-554.
15

Zakian, V. A. (1995). Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.

Patentansprüche

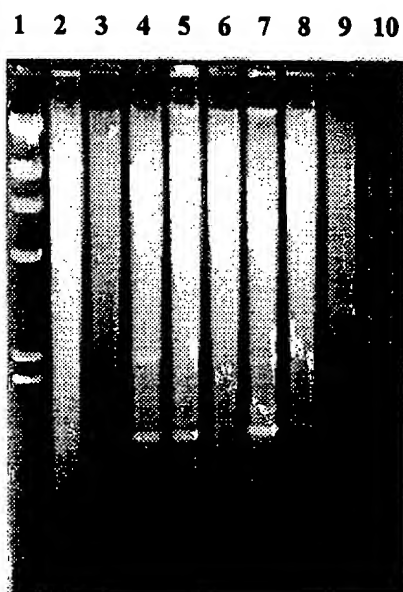
1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen
Telomerase-Untereinheit.
2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um
Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser
Sequenzen handelt.
3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um
die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen
katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder
um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.
4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der
Ansprüche 1 bis 3.
5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß
es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide
oder Proteine kodieren.
6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.
7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem
der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.
8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vek-
toren gemäß einem der Ansprüche 4 bis 6.

9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:
- 5 A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,
- B. Messung des Substanzeffektes auf die Reportergenexpression.
- 10
10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screent.
- 15
11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.
- 20
12. Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:
- A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält
- 25 mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

- C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

Fig. 1

A



B

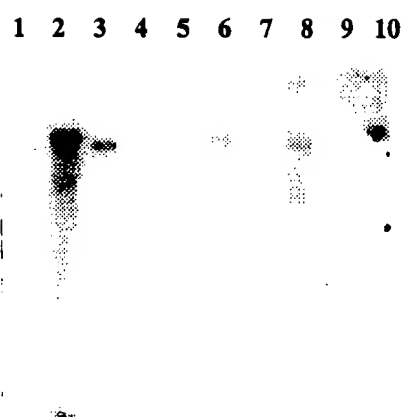


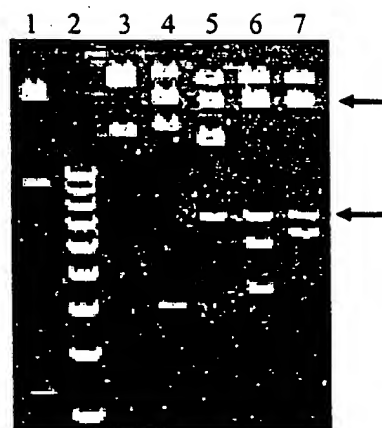
Fig. 2

Fig. 3

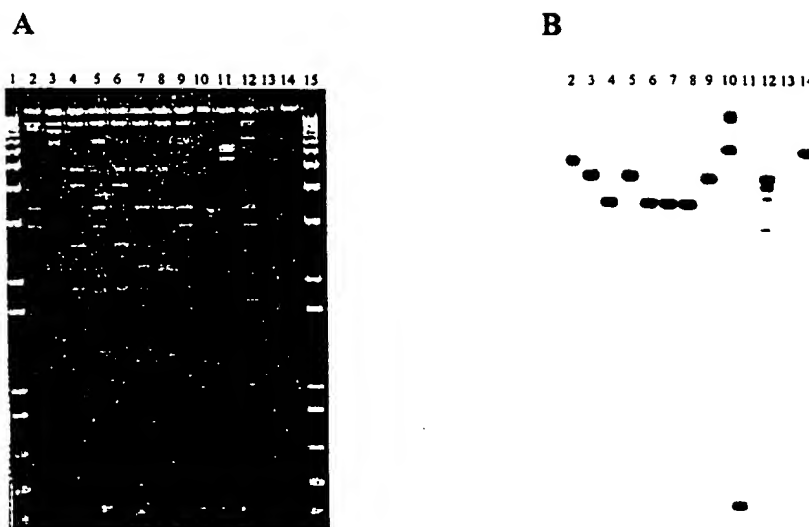


Fig. 4

GAGCTCTGAA	CCGTGGAAAC	GAACATGACC	CTTGCCCTGCC	TGCTTCCCTG	GGTGGGTCAA	GGGTAATGAA	70
TGGGTGTGCA	GGAAATGGCC	ATGTAAATTA	CACGACTCTG	CTGATGGGGA	CCGTTCCCTC	CATCATATT	140
CATCTTCACC	CCCAAGGACT	GAATGATTCC	AGCAACTTCT	TGGGTGTGTA	CAAGCCATGA	CAAAACTCAG	210
TACAAACACC	ACTCTTTTAC	TAGGCCCCACA	GAGCACGGGC	CACACCCCTG	ATATATTAA	AGTCCAGGAG	280
AGATGAGGCT	GCTTTCAGCC	ACCAGGCTGG	GGTGACAACA	GCGGCTGAAC	AGTCTGTTC	CTAGACTAG	350
TAGACCCCTGG	CAGGCACTCC	CCCAAAATCT	AGGGCCTGGT	TGCTGCTTCC	CGAGGGCGCC	ATCTGCCCTG	420
GAGACTCAGC	CTGGGGTGCC	ACACTGAGGC	CAGCCCTGTC	TCCACACCCT	CCGCCCTCCG	GCCTCAGCTT	490
CTCCAGCAGC	TTCTTAAACC	CTGGGTGGGC	CGTGTTCAG	CGCTACTGTC	TCACCTGTCC	CAGTGTGTCT	560
TGCTCAGCG	ACGTAGCTCG	CACGGTTCCT	CCTCACATGG	GGTGTCTGTC	TCCTTCCCCA	ACACTCACAT	630
GCCTTGAAGG	GAGGAGATTG	TGCGCCTCCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGGCCCCC	700
GATGCAGGTT	CCTGGCGTCC	GGCTGCACGC	TGACCTCCAT	TTCCAGGCGC	TCCCCGTCTC	CTGTCACTCTG	770
CCGGGGCCTG	CCGGTGTGTT	CTTCTGTTTC	TGTGCTCCTT	TCCACGTCCA	GCTGCGTGTG	TCTCTGCCCG	840
TAGGGTCTC	GGGGTTTTTA	TAGGCATAGG	ACGGGGGCGT	GGTGGGCCAG	GGCGCTCTTG	GGAAATGCAA	910
CATTGGGGTG	TGAAAGTAGG	AGTGCCCTGTC	CTCACCTAGG	TCCACGGGCA	CAGGCCTGGG	GATGGAGCCC	980
CCCGCAGGGA	CCCGCCCTTC	TCTGCCCAGC	ACTTTCCTGC	CCCCCTCCCT	CTGGAACACA	GAGTGGCAGT	1050
TTCCACAAGC	ACTAAGCATC	CTCTTCCCAA	AAGACCCAGC	ATTGGCACCC	CTGGACATT	GCCCCACAGC	1120
CCTGGGAATT	CACGTGACTA	CGCACATCTC	GTACACACTC	CCGTCCACGA	CCGACCCCTG	CTGTTTTATT	1190
TTAATAGCTA	CAAAGCAGGG	AAATCCCTGC	TAAATGTCC	TTTAACAAAC	TGGTTAAACA	AACGGGTCCA	1260
TCCGCACGGT	GGACAGTTCC	TCACAGTGAA	GAGGAACATG	CCGTTTATAA	AGCCTGCAGG	CATCTCAAGG	1330
GAATTACGCT	GAGTCAAAAC	TGCCACCTCC	ATGGGATACG	TACGCAACAT	GCTCAAAAAG	AAAGAATTTC	1400
ACCCCATGGC	AGGGGAGTGG	TTAGGGGGGT	TAAGGACGGT	GGGGGCGGCA	GCTGGGGGCT	ACTGCACGCA	1470
CCTTTTACTA	AAGCCAGTTT	CCTGGTCTCG	ATGGTATTGG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	1540
TGGGAGTGGG	GGAAACCCGA	GGCTGTGCCA	TCTTTGCCAT	GCCCGAGTGT	CCTGGGCGAG	ATAATGCTCT	1610
AGAGATGCCC	ACGTCTCTGAT	TCCCCCAAAC	CTGTGGACAG	AACCCGCCCG	GCCCCAGGGC	CTTTGCAGGT	1680
GTGATCTCCG	TGAGGACCCCT	GAGGTCTGGG	ATCCTTCGGG	ACTACCTGCA	GGCCCCGAAA	GTAAATCCAGG	1750
GTTTCTGGGA	AGAGCCGGGC	AGGAGGGTCA	GAGGGGGGCA	GCCTCAGGAC	GATGGAGGCA	GTCACTCTGA	1820
GGCTGAAAAG	GGAGGGAGGG	CCTCGAGCCC	AGGCTTGCAA	GCGCCTCCAG	AAGCTGGAAA	AAGCGGGGAA	1890
GGGACCCCTCC	ACGGAGCCTG	CAGCAGGAAG	GCACGGCTGG	CCCTTAGCCC	ACCAGGGCCC	ATCGTGGACC	1960
TCCGGCCCTCC	GTGCCATAGG	AGGGCACTCG	CGCTGCCCTT	CTAGCATGAA	GTGTGTGGGG	ATTTCAGAAA	2030
GCAACAGGAA	ACCCATGCAC	TGTGAATCTA	GGATTATTTC	AAAACAAAGG	TTTACAGAAA	CATCCAAGGA	2100
CAGGGCTGAA	GTCCCTCCGG	GCAAGGGCAG	GGCAGGCACG	AGTGATTTTA	TTTAGCTATT	TTATTTTATT	2170
TACTTACTTT	CTGAGACAGA	GTTATGCTCT	TGTTGCCCAG	GCTGGAGTGC	AGCGGCATGA	TCTTGGCTCA	2240
CTGCAACCTC	CGTCTCCTGG	GTTCAAGCAA	TTCTCGTGCC	TCAGCCTCCC	AAGTAGCTGG	GATTTCAGGC	2310
GTGCACCACC	ACACCCGGCT	AATTTTGTAT	TTTTAGTAGA	GATGGGCTTT	CACCATGTTG	GTCAAGCTGA	2380
TCTCAAAATC	CTGACCTCAG	GTGATCCGCC	CACCTCAGCC	TCCCAAAGTG	CTGGGATTAC	AGGCATGAGC	2450
CACGTCACCT	GGCCTATTTA	ACCATTTTAA	AACTTCCCTG	GGCTCAAGTC	ACACCCACTG	GTAAAGGAGT	2520
CATGGAGTTT	AATTTCCCTT	TTACTCAGGA	GTTACCCCTC	TTTGATATTT	TCTGTAGATT	TCTGTAGACT	2590
GGGGATACAC	CGTCTCTTGA	CATATTACAC	GTTTCTGTGA	CCACCTGTTA	TCCCATGGGA	CCCCTGCAG	2660
GGGCAGCTGG	GAGGCTGCAG	GCTTCAGGTC	CCAGTGGGGT	TGCCATCTGC	CAGTAGAAAC	CTGATGTAGA	2730
ATCAGCGGCG	AAGTGTGGAG	ACTGTCTCTG	ATCTCAATGT	CTCAGTGTGT	GCTGAATAAT	GTGAATAATTA	2800
AAGTCCATCC	CTCCTACTCT	ACTGGGATTG	AGCCCCCTCC	CTATCCCCCC	CCAGGGGCGG	AGGAGTTTCT	2870
CTCACTCTCT	TGGAGGAAGG	AATGATACTT	TGTTATTTT	CACTGCTGGT	ACTGAATCCA	CTGTTTCATT	2940
TGTTGGTTTG	TTTGTTTTGT	TTTGAGAGGC	GGTTTCACTC	TTGTTGCTCA	GGCTGGAGGG	AGTCCAATGG	3010
CGCGATCTTG	GCTTACTGCA	GCCTCTGCCT	CCCAGGTTCA	AGTGATTCTC	CTGCTTCCGC	CTCCCATTTG	3080
GCTGGGATTA	CAGGCACCCG	CCACCATGCC	CAGCTAATTT	TTTGATTTT	TAGTAGAGAC	GGGGGTGGGT	3150
GGGGTTTACC	ATGTTGGCCA	GGCTGGTCTC	GAACCTCTGA	CCTCAGATGA	TCCACCTGCC	TCTGCCCTCT	3220
AAAGTGCCTG	GATTACAGGT	GTGAGCCACC	ATGCCACGCT	CAGAATTTAC	TCTGTTTAGA	AACATCTGGG	3290
TCTGAGGTAG	GAAAGCTCACC	CCACTCAAGT	GTTGTGGTGT	TTTAAGCCAA	TGATAGAAAT	TTTTTATTGT	3360
TGTTAGAAAC	CTCTTGATGT	TTTACACTGT	GATGACTAAG	ACATCATCAG	CTTTTCAAAG	ACACACTAAC	3430
TGCACCCATA	ATACTGGGGT	GTCTTCTGGG	TATCAGCAAT	CTTCATTGAA	TGCCGGGAGG	CGTTTCCCTG	3500
CCATGCACAT	GGTGTTAATT	ACTCCAGCAT	AATCTTCTGC	TTCCATTTC	TCTCTTCCCT	CTTTTAAAAAT	3570
TGTGTTTTCT	ATGTTGGCTT	CTCTGCAGAG	AACCAGTGTA	AGCTACAAC	TAACTTTTGT	TGGAACAAAT	3640
TTTCCAAACC	GCCCTTTTGC	CCTAGTGGCA	GAGACAATTC	ACAAACACAG	CCCTTTAAAA	AGGCTTAGGG	3710
ATCACTAAGG	GGATTTCATG	AAGAGCGACC	TGTAATCCTA	AGTATTTACA	AGACGAGGCT	AACCTCCAGC	3780
GAGCGTGACA	GCCACGGGAG	GGTGCGAGGC	CTGTTCAAAT	GCTAGCTCCA	TAAATAAAGC	AATTTCTCTC	3850
GGCAGTTTCT	GAAAGTAGGA	AAGGTTACAT	TTAAGGTTGC	GTTTGTTAGC	ATTTCAAGT	TTGCCGACCT	3920
CAGCTACAGC	ATCCCTGCAA	GGCCTCGGGA	GACCCAGAAG	TTTCTCGCCC	CCTTAGATCC	AAACTTGGAG	3990
AACCCGGAGT	CTGGATTCTT	GGGAAGTCTT	CAGCTGTCTT	GCGGTTGTGC	CGGGCCCCCA	GGTCTGGAGG	4060
GGACCACTGG	CCGTGTGGCT	TCTACTGCTG	GGCTGGAAGT	CGGGCTCTCT	AGCTCTGCAG	TCCGAGGCTT	4130
GGAGCCAGGT	GCCTGGACCC	CGAGGCTGCC	CTCCACCTTG	TGCGGGCGGG	ATGTGACCAG	ATGTTGGCCT	4200
CATCTGCCAG	ACAGAGTGCC	GGGGCCCAAG	GTCAAGGCGG	TTGTGGCTGG	TGTGAGGCGC	CCGGTCCGCG	4270
GCCAGCAGGA	GCGCCTGGCT	CCATTTCACA	CCCTTCTCTG	ACGGGACCGC	CCCGGTGGGT	GATTAAACAGA	4340
TTTGGGGTGG	TTTGCTCATG	TGGGGGAGCC	CTCGCCGCTT	GAGAACCTGC	AAAGAGAAAT	GACGGGCGCT	4410
TGTCAAGGAG	CCCAAGTCCG	GGGGAAGTGT	TGCAGGGAGG	CACTCCGGGA	GGTCCCGCGT	CCCCGTCCAG	4480
GGAGCAATGC	GTCTCTGGGT	TGCTCCCCAG	CCGCGTCTAC	GCGCCTCCGT	CCTCCCTTTC	ACGTCGGGCA	4550
TTGCTGGTGC	CCGGAGCCCG	ACGCCCCCGG	TCCGGACCTG	GAGGCAGCCC	TGGGTCTCCG	GATCAGGCCA	4620
GCGGCCAAAG	GGTGGCCGCA	CGCACCTGTT	CCCAGGGCCT	CCACATCATG	GCCCTTCCCT	CGGGTTACCC	4690

Fig. 4 (Fortsetzung)

```
CACAGCCTAG GCCGATTCGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA 4760
GCGGCGCGCG GCGGGGAAG CGCGGCCAG ACCCCCGGGT CCGCCCGGAG CAGCTGCGCT GTCGGGGCCA 4830
GGCCGGGCTC CCAGTGGATT CGCGGGCACA GACGCCAGG ACCGCGCTCC CCACGTGGCG GAGGGACTGG 4900
GGACCCGGGC ACCCGTCCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCCCGTCCC 4970
GACCCCTCCC GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCT CCCCTTCCTT TCCCGGGCCC 5040
CGCCCTCTCC TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCGC ACGTGGGAAG CCCTGGCCCC 5110
GGCCACCCC GCGATG 5126
```

Fig. 5

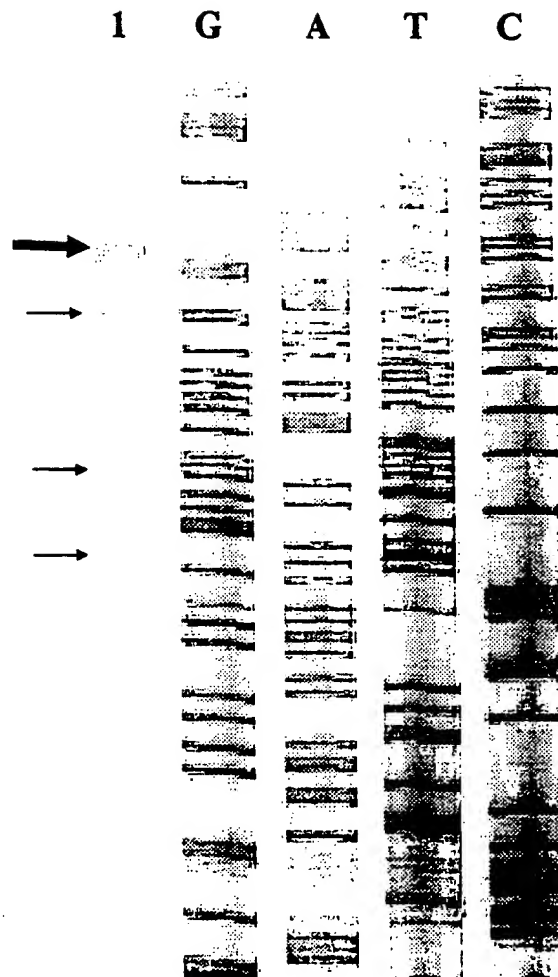
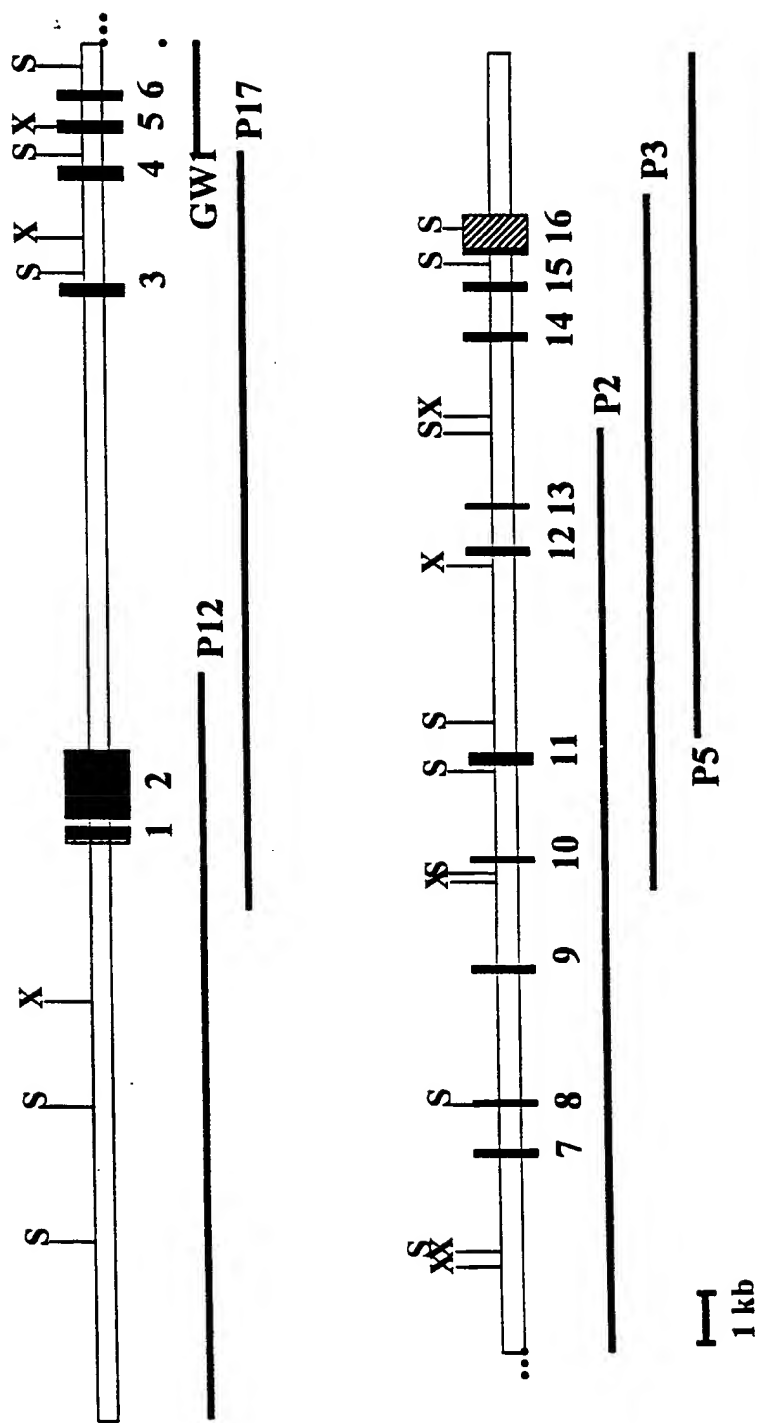


Fig. 6

GTTCAGGCA	GCGCTGCGTC	CTGCTGCGCA	CGTGGGAAGC	CCTGGCCCCG	GCCACCCCCG	CGATGCCCGG	70
CGCTCCCCG	TGCCGAGCCG	TGCGCTCCCT	GCTGCGCAGC	CACTACCCGG	AGGTGCTGCC	GCTGGCCAC	140
TTCTGTGCGG	GCCTGGGGCC	CCAGGGCTGG	CGGCTGGTGC	AGCGCGGGGA	CCCGCGGGCT	TTCCGCGCGC	210
TGGTGGCCCA	GTGCTGCTG	TGCGTGCCCT	GGGACGCACG	GCCGCCCCCC	GCCGCCCCCT	CCTTCCGCCA	280
GGTGTCTTGC	CTGAAGGAGC	TGGTGGCCCG	AGTGTGCGAG	AGGCTGTGCG	AGCGCGGGCG	GAAGAAGCTG	350
CTGGCCTTCG	GCTTCGCGCT	GCTGGACGGG	GCCCCGCGGG	GCCCCCCCCG	GGCCTTCACC	ACCAGCGTGC	420
GCAGCTACCT	GCCCAACACG	GTGACCGACG	CACCTGCGGG	GAGCGGGGCG	TGGGGGCTGC	TGCTGCGCCG	490
CGTGGGCGAC	GACGTGCTGG	TTACCTGTCT	GGCAGCGTGC	GCGCTCTTTG	TGCTGGTGGC	TCCCAGCTGC	560
GCCTACCAGG	TGTGCGGGCC	GCCGCTGTAT	CAGCTCGGCG	CTGCCACTCA	GGCCCCGGCC	CCGCCACACG	630
CTAGTGGACC	CCGAAGGCGT	CTGGGATGCG	AACGGGCTTG	GAACCATAGC	GTGAGGGAGG	CCGGGTGCC	700
CTGGGGCTTG	CCAGCCCCGG	GTGCGAGGAG	GCGCGGGGCG	AGTGCCAGCC	GAAGTCTGCC	GTTGCCCAAG	770
AGGCCCCAGG	GTGGCGCTGC	CCCTGAGCCG	GAGCGGACGC	CCGTGGGCA	GGGGTCTTGG	GCCCCACCCG	840
GCAGGACCGG	TGGACCGAGT	GACCGTGGTT	TCTGTGTGGT	GTACCTGCC	AGACCCGCGG	AAGAAGCCAC	910
CTCTTTGAG	GGTGCCTCT	CTGGCAGCGG	CCACTCCAC	CCATCCGTGG	GCCGCGAGCA	CCAGCGGGCG	980
CCCCCATCCA	CATCGCGGCC	ACCACGTCCC	TGGGACACGC	CTTGTCCTCC	GGTGTACGCC	GAGACCAAGC	1050
ACTTCTCTTA	CTCTCTCAGC	GACAAGGAGC	AGCTGCGGCC	CTCCTTCCTA	CTCAGCTCTC	TGAGGCCCGG	1120
CCTGACTGGC	GCTCGGAGGC	TCGTGGAGAC	CATCTTTCTG	GGTTCAGGC	CCTGGATGCC	AGGGAATCCC	1190
GCGAGGTTGC	CCCCCTGCC	CCAGCGCTAC	TGGCAATGCG	GGCCCCGTGT	TCTGGAGCTG	CTTGGGAACC	1260
ACGCGCAGTG	CCCCCTACGG	GTGCTCTCTA	AGACGCACTG	CCCCGTGCGA	GCTGCGGCTC	CCCCAGCAGC	1330
CGGTGTCTGT	GCCCCGGAGA	AGCCCCAGGG	CTCTGTGGCG	GCCCCCGAGG	AGGAGGACAC	AGACCCCCGT	1400
GCCCTGGTGC	AGCTGCTCCG	CCAGCAGACG	AGCCCCGTGC	AGGTGTACGG	CTTCGTGCGG	GCTGCTGCTG	1470
CCCCGGCTGG	GCCCCCAGGG	CTCTGGGGCT	CCAGGACCAA	GGAACGCGCG	TTCTCTCAGG	ACACCAAGAA	1540
GTTCATCTCC	CTGGGGAAGC	ATGCCAAGCT	CTCGCTGCAG	GAGCTGACGT	GGAAGATGAG	CGTCCGGGAC	1610
TGCGCTTGGC	TGCGCAGGAG	CCCAGGGGTT	GGCTGTGTTT	CGGCGCGAGA	GCACCTCTCT	CGTGAGGAGA	1680
TCCTGGCCAA	GTTCTGCGAC	TGGCTGATGA	GTGTGTACGT	CGTGGAGCTG	CTCAGGTCTT	TCTTTTATGT	1750
CACGGAGACC	ACGTTTCAAA	AGAACAGGCT	CTTTTCTTAC	CGGAAGAGTG	TCTGGAGCAA	GTGCGAAAGC	1820
ATTGGAATCA	GACAGCACTT	GAAGAGGGTG	CAGCTGCGGG	AGCTGTGCGA	AGCAGAGGTC	AGGCAGCATC	1890
GGGAAGCCAG	GCCCCGCCCT	CTGACGTCCA	GACTCCGCTT	CATCCCCAAG	CCTGACGGGC	TGCGGCCGAT	1960
TGTGAACATG	GACTACGTGG	TGGGAGCCAG	AACGTTCCGC	AGAGAAAAGA	GGGCGGAGCG	GCTCACCTCG	2030
AGGTGGAAGG	CACCTGTTCA	CGTCTCAAC	TACGAGCGGG	CGCGCGCCCC	CGGCTCTCTG	GCGGCTCTCT	2100
TGCTGGGGCT	GGACGATATC	CACAGGGGCT	GGCGCACCTT	CGTGTGCGGT	GTGCGGGCCC	AGGACCCGCG	2170
GCCTGAGCTG	TACTTTGTCA	AGGTGGATGT	GACGGGCGCG	TACGACACCA	TCCCCCAGGA	CAGGCTCAGC	2240
GAGGTGATCG	CCAGCATCAT	CAAAACCCAG	AACACGTACT	CGGTGCGTGC	GTATGCCGTG	GTCCAGAAGG	2310
CCGCCCATGG	GCACGTCCCG	AAGGCCTTCA	AGAGCCACGT	CTCTACCTTG	ACAGACCTCC	AGCCGTACAT	2380
GCGCAGTTTC	GTGGCTCACC	TGCAGGAGAC	CAGCCCCGTC	AGGGATGCCG	TGCTCATCGA	GCAGAGCTCC	2450
TCCCTGAATG	AGGCCAGCAG	TGGCTCTTTC	TACGCTTTCAT	TACGCTTTCAT	GTGCCACACC	CCGCTGACCA	2520
TCAGGGGCAA	GTCCCTACGT	CAGTGCCAGG	GGATCCCGCA	GGGCTCCATC	CTCTCCACGC	TGCTCTGCAG	2590
CCTGTGCTAC	GGCGACATGG	AGAACAAAGT	GTTTGGGGGG	ATTGCGCGGG	ACGGGCTGCT	CCTGCGTTTG	2660
GTGGATGATT	TCTTGTGGGT	GACACCTCAC	CTCACCCACG	CGAAACCTT	CCTCAGGACC	CTGCTCCGAG	2730
GTGTCCCTGA	GTAAGGCTGC	GTGGTGAAGT	TGCGGAAGAC	AGTGGTGAAC	TTCCCTGTAG	AAGACGAGGC	2800
CCTGGGTGGC	ACGGCTTTTG	TTCAAGTGCC	GGCCACCGGC	CTATTCCCTT	GGTGGCGGCT	GCTGCTGGAT	2870
ACCCGGACCC	TGGAGGTGCA	GAGCGACTAC	TCCAGCTATG	CCCGGACCTC	CATCAGAGCC	AGTCTCACCT	2940
TCAACCGCGG	CTTCAAGGCT	GGGAGGAACA	TGCGTCCGAA	ACTCTTTGGG	GTCTTGGCGG	TGAAGTGTCA	3010
CAGCCTGTTT	CTGGATTTCG	AGGTGAACAG	CCTCCAGACG	GTGTGCACCA	ACATCTACAA	GATCCTCTCT	3080
CTGCAGGCGT	ACAGGTTTCA	CGCATGTGTG	CTGCAGCTCC	CATTTTCATC	GCAAGTTTGG	AAGAACCCTA	3150
CATTTTTCCT	GCGCGTCAAT	TCTGACACGG	CCTCCCTCTG	CTACTCCATC	CTGAAAGCCA	AGAACGCGAG	3220
GATGTGCTGT	GGGGCCAAAG	GCGCGCGCGG	CCCTCTGCCC	TCCGAGGCGG	TGCAAGTGGT	GTGCCACCAA	3290
GCATTCCTGC	TCAAGCTGAC	TGCACACCGT	GTCACCTACG	TGCCACTCCT	GGGGTCACTC	AGGACAGCCC	3360
AGACCGAGCT	GAGTCCGAAG	CTCCCGGGGA	CGACGCTGAC	TGCCCTGGAG	GCCGCGACCA	ACCCGGCACT	3430
GCCCTCAGAC	TTCAAGACCA	TCCTGGACTG	ATGGCCACCC	GCCCACAGCC	AGGCCGAGAG	CAGACACCAG	3500
CAGCCTGTGC	ACGCCCGGCT	CTACGTCCCA	GGGAGGGAGG	GGCGGCCCCC	ACCCAGGCCC	GCACCGCTGG	3570
GAGTCTGAGG	CCTGAGTGAG	TGTTTGGCCG	AGGCTGCGAT	GTCCGGCTGA	AGGCTGAGTG	TCCGGCTGAG	3640
GCCTGAGCGA	GTGTCCAGCC	AAGGGCTGAG	TGTCCAGCAC	ACCTGCGGTC	TTCACTTCCC	CACAGGCTGG	3710
CGCTCGGCTC	CACCCAGGGG	CCAGCTTTTC	CTCACCAGGA	GCCCCGGCTT	CACCTCCCCAC	ATAGGAATAG	3780
TCCATCCCCA	GATTGCGCAT	TGTTTACCCC	TGCCCCCTGC	CTCCTTTGCC	TTCCACCCCC	ACCATCCAGG	3850
TGGAGACCTT	GAGAAGGACC	CTGGGAGCTC	TGGGAATTTG	GAGTGACCAA	AGGTGTGCCC	TGTACACAGG	3920
CGAGGACCTT	GCACCTGGAT	GGGGTCCCTT	GTGGGTCAAA	TTGGGGGGAG	GTGCTGTGGG	AGTAAATATC	3990
TGAATATATG	AGTTTTTTCAG	TTTTTGAATA	AAAAAAAAAA	AAAAAAAAAA	AA		4042

ERSATZBLATT (REGEL 26)

Fig. 7



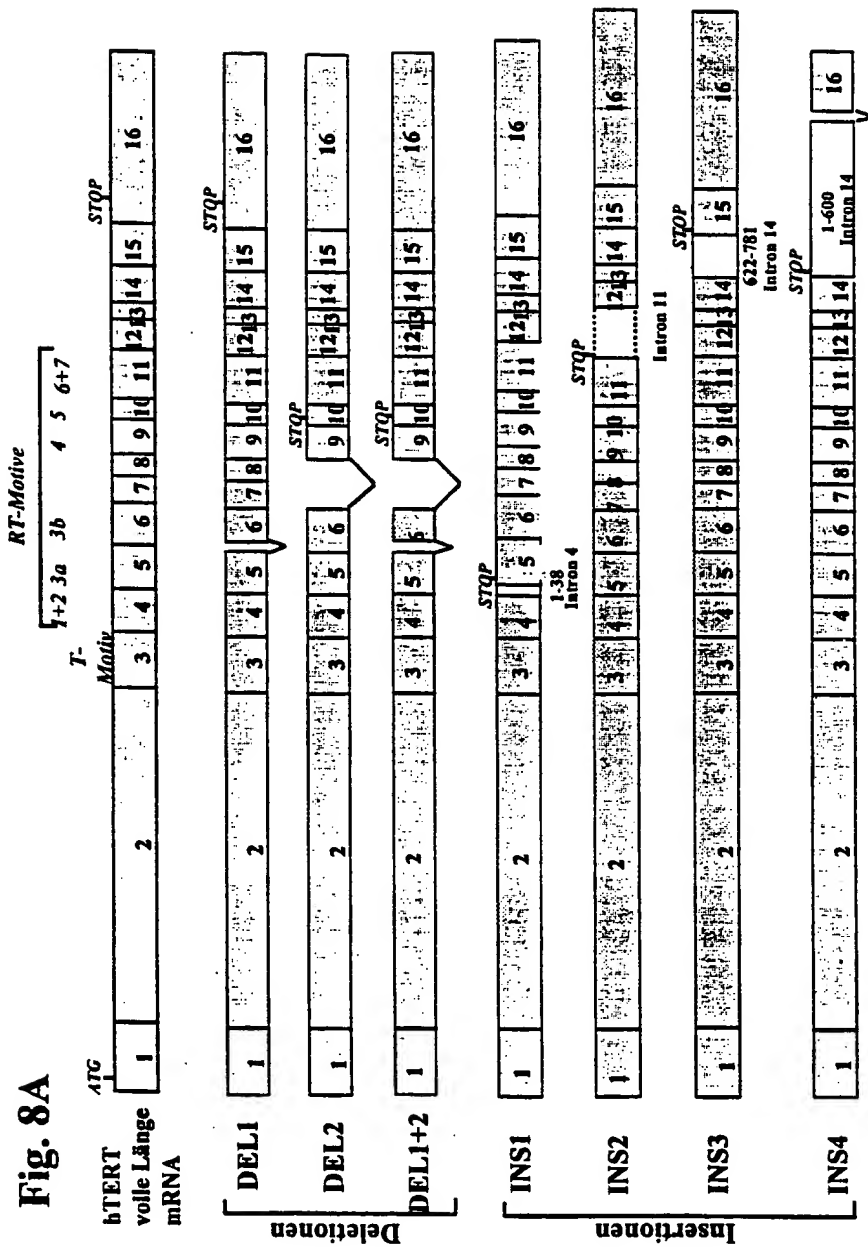


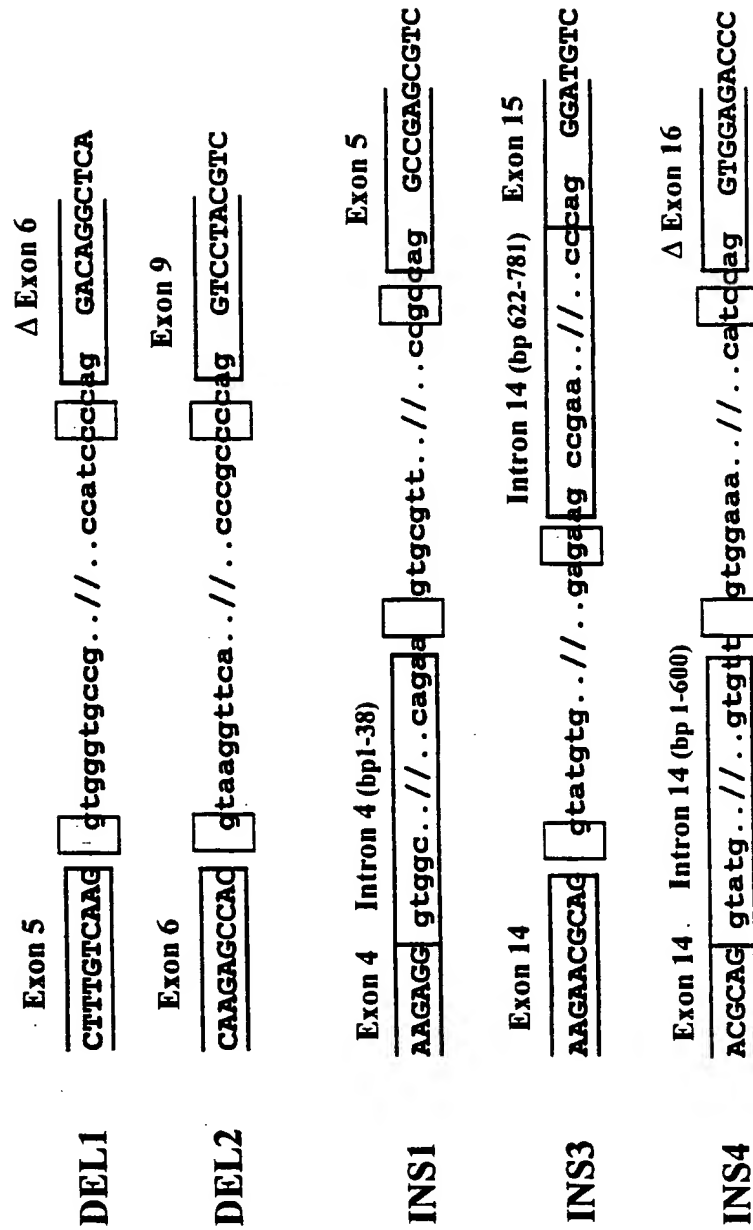
Fig. 8B

Fig. 9

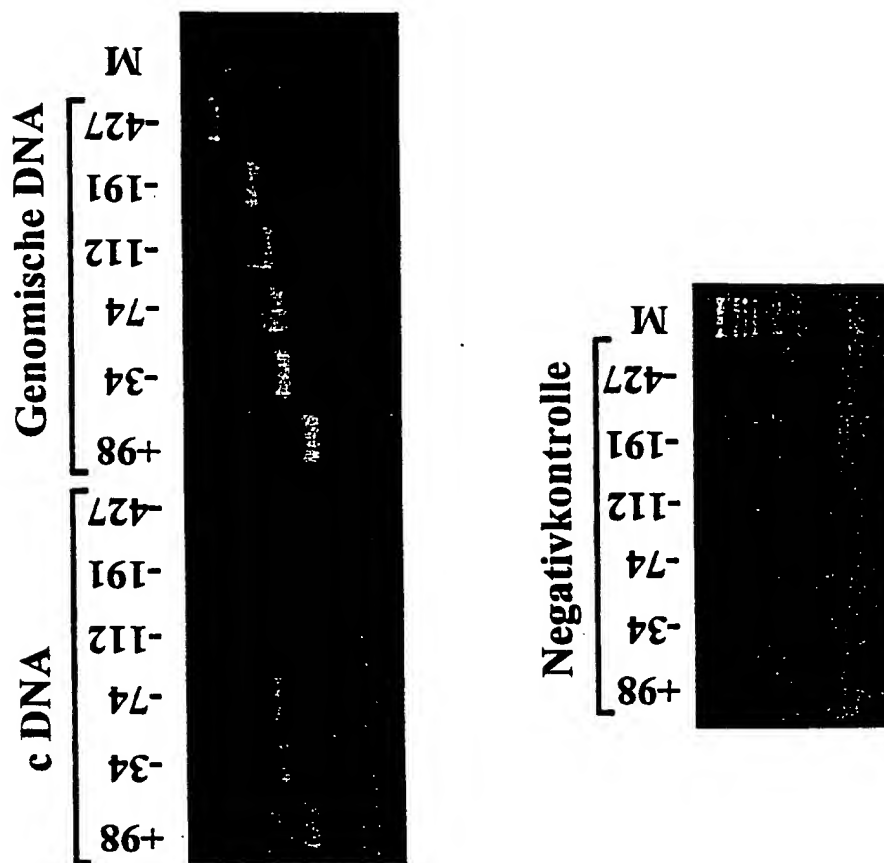


Fig. 10

ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA -11204
 ATGAGACCCT GTCTCAAAAA AAAAAAAAAA AATTGAAATA ATATAAAGCA TCTTCTCTGG CCACAGTGGG -11134
 AAAAAACCAG AAATCAACAA CAAGAGGAAT TTTGAAACT ATACAAACAC ATGAAAATTA AACAAATATAC -11064
 TTCTGAATGA CCAGTGAGTC AATGAAGAAA TTAAGAAAGG AATTGAAAAA TTTATTTAAG CAAATGATAA -10994
 CGGAACATA ACCTCTCAAA ACCCAGCGTA TACAGCAAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA -10924
 AGCAGCTACA TCAAAAAAGT AGAAAAGCCA GGCGCAGTGG CTCATGCCTG TAATCCCAGC ACTTTGGGAG -10854
 GCCAAGGCGG GCAGATCGCC TGAGGTCAGG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTGC -10784
 CTACTAAAAA TACAAAATTA GCTGGGCATG GTGGCACATG CCTGTAATCC CAGCTACTCG GGAGGCTGAG -10714
 GCAGGATAAC CGCTTGAACC CAGGAGGTGG AGGTTGCGGT GAGCCGGGAT TGCGCCATTG GACTCCAGCC -10644
 TGGGTAACAA GAGTGAAACC CTGTCTCAAG AAAAAAAAAA AAGTAGAAAA ACTTAAAAAT ACAACCTAAT -10574
 GATGCACCTT AAAGAAGTAG AAAAGCAAGA GCAAACTAAA CCTAAAATTG GTAAAAAGAA AGAAATAATA -10504
 AAGATCAGAG CAGAAATAAA TGAACTGAA AGATAACAA ACAAAAGATC AACAAAATTA AAGTTGGTT -10434
 TTTTGAAAG ATAAACAAAA TTGACAAACC TTTGCCAGA CTAAGAAAAA AGGAAGAAG ACCTAAATAA -10364
 ATAAAGTCAG AGATGAAAAA AGAGACATTA CACTGATAC CACAGAAAT CAAAGGATCA CTAGAGGCTA -10294
 CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCCTAGA TGCATACAAC -10224
 CTACCAAGT TGAACCATGA AGAAATCCAA AGCCCAACA GACCAATAAC AATAATGGGA TTAAGCCAT -10154
 AATAAAGT CTCCTAGCAA AGAGAAGCCC AGGACCCAAT GGCTTCCCTG CTGGATTTTA CCAATCATTT -10084
 AAAGAAGAA GAATTCCAAT CTTACTCAA CTATTCTGAA AAATAGAGGA AAGAATACTT CCAAACTCAT -10014
 TCTACATGGC CAGTATTACC CTGATTCCAA AACCAGACAA AAACACATCA AAAACAAAAA AACAAAAAAA -9944
 CAGAAAGAAA GAAACTACA GGCCAATATC CCTGATGAAT ACTGATACAA AAATCCTCAA CAAAACATA -9874
 GCAAAACAAA TTAACAACA CCTTCGAAAG ATCATTCTAT GTGATCAAGT GGGATTTATT CCAGGGATGG -9804
 AAGGATGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAAACTA -9734
 TATGATTATT TCACTTTATG CAGAAAAAGC ATTTGATAAA ATTCTGCACC CTTCTATGTA AAAACCTCA -9664
 AAAAACCCAG TATACAAGAA ACATACAGGC CAGGCACAGT GGCTCACACC TGCGATCCCA GCACTCTGGG -9594
 AGGCCAAGGT GGGATGATTG CTTGGGCCCA GGAGTTTGAG ACTAGCCTGG GCAACAAAAT GAGACCTGGT -9524
 CTACAAAAA CTTTTTAAA AAATTAGCCA GGCGATGAG CATATGCCTG TAGTCCCAGC TAGTCTGGAG -9454
 GCTGAGTGG GAGAATCACT TAAGCCTAGG AGGTCGAGGC TGCACTGAGC CATGAACATG TCACTGTACT -9384
 CCAGCCTAGA CAACAGAACA AGACCCCACT GAATAAGAAG AAGGAGAAG AGAAGGGAGA AGGGAGGGAG -9314
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAAGTGG AAGGGGAAGG GGAAGGGAAA GAGGAAGAAG -9244
 AAGAAACATA TTCAACATA ATAAAGCCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAAGTGA -9174
 AGCCTTTCCT CTAAGATCTG GAAAATGACA AGGGCCCACT TTCACCACTG TGATTCAACA TAGTACTAGA -9104
 AGTCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAAA GGCATCCAAA CTGGAAGGA AGAAGTCAAA -9034
 TTATCTGTT TGCAGATGAT ATGATCTTAT ATCTGAAAA GACTTAAGAC ACCACTAAAA AACTATTAGA -8964
 GCTGAAATTT GGTACAGCAG GATACAAAAA CAATGTACAA AAATCAGTAG TATTTCTATA TTCCAACAGC -8894
 AAACAATCTG AAAAAAGAAC CAAAAAGCA GCTACAAATA AAATTAACA GCTAGGAATT AACCAAGAA -8824
 GTGAAAGAT TCTACAATGA AAATAATAA ATGTTGATAA AAGAAATTGA AGAGGGCACA AAAAAAGAA -8754
 AGATAATCCA TGTTCATAGA TTGGAAGAAT AAATACTGTT AAATGTCCA TACTACCCA AGCAATTTAC -8684
 AAATTCATG CAATCCCTAT TAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTCTAAGAT -8614
 TTGTACAGAA CCACAAAAGA CCCAGAAATG CCAAGCTAT CCTGACCAA AAGAACAAA CTGGAAGCAT -8544
 CACATTAACCT GACTTCAAT TATATACAA AGCTATAGTA ACCCAAATA CATGGTACTG GCATAAAAC -8474
 AGATGAGACA TGGACCAGAG GAACAGAATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAATCA -8404
 TTTTGTACAA AGGTGCCAAG AACATACCTT GGGGAAAAA TAATCTCTC AATAAATGGT GCTGGAGGAA -8334
 CTGGATATCC ATATGCAAAA TAACAATCT AGAATCTGT CTCTCACCAT ATACAAAAGC AAATCAAAAT -8264
 GGATGAAAGG CTTAAATCTA AAACCTCAAA CTTTGCAACT ACTAAAAGAA AACACCCGGG AAATCTTCCA -8194
 GGACATTGGA GTGGGCAAG ACTTCTTGAG TAATCCCTG CAGGCACAG CAACCAAGC AAAACAGAC -8124
 AAATGGGATC ATATCAAGTT AAAAGCTTC TGCCCGACAA AGGAAACAAT CAACAAAGAG AAGAGACAAC -8054
 CCACAGAATG GGAGAATATA TTTGCAAACT ATTCATCTAA CAAGGAATTA ATAACCAGTA TATATAAGGA -7984
 GCTCAAACTA CTCTATAAGA AAAACACCTA ATAAGCTGAT TTTCAAAAAT AAGCAAAAGA TCTGGGTAGA -7914
 CATTTCTCAA AATAAGTCAT ACAAATGGCA AACAGGCATC TGAAAATGTG CTCAACACCA CTGATCATCA -7844
 GAGAAATGCA AATCAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAAA ATGGCTTTTA TTCAAAAGAC -7774
 AGGCAATAAC AAATGCCAGT GAGGATGTGG ATAAAAGGAA ACCCTTGGAC ACTGTTGGTG GGAATGGAAA -7704
 TTGCTACCAC TATGGAGAAC AGTTTGAAAG TTCCTCAAAA AACTAAAAAT AAAGCTACCA TACAGCAATC -7634
 CCATTGCTAG GTATATACTC CAAAAAGGG AATCAGTGA TCAACAAGCT ATCTCCACTC CCACATTTAC -7564
 TGCAGCACTG TTCATAGCAG CCAAGGTTTG GAAGCAACCT CAGTGTCCAT CAACAGACGA ATGGAAAAAG -7494
 AAAATGTGGT GCACATACAC AATGGAGTAC TACGCAGCCA TAAAAAGAA TGAGATCCTG TCAGTTGCAA -7424
 CAGCATGGG GGCATGGTC AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAA GTTTCATGTT -7354
 CTCCTTACT TGTGGGAGCA AAAATTAATA CAATTGCAT AGAAATAGAG GAGAATGGTG GTTCTAGAGG -7284
 GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTTATTG TATGTTTAA AATRACTAAA AGAGTATAAT -7214
 TGGGTTGTT GTAACACAAA GAAAGGATAA ATGCTTGAAG GTGACAGATA CCCATTATC CTTGATGTA -7144
 TTATTACACA TTGTATGCCT GTATCAAAAT ATCTCATGTA TGCTATAGAT ATAAACCCTA CTATATTA -7074
 AATTAATAAT TTAATGGCCA GGCACGGTGG CTCATGTCCG TAATCCAGC ACTTTGGGAG GCCGAGGCGG -7004
 GTGGATCACC TGAGGTGAGG AGTTTGAAC CAGTCTGGCC ACCATGATGA AACCTGTCT CTAATAAAGA -6934
 TACAAAAAT AGCCAGGCGT GGTGGCAGT ACCTGTAGT CCAACTACTC AGGAGGCTGA GACGAGAGAA -6864
 TTGCTTGAAC CTGGGAGGCG GAGGTTGCAG TGAGCCGAGA TCATGCCACT GCACTGCAGC TTGGGTGACA -6794
 GAGCAAGACT CCATCTCAA ACAAACAAA AAAAAAGAG ATTAATAATG TAATTTTAT GTACCGTATA -6724
 AATATATACT CTACTATATT AGAAGTTAAA AATTAACA ATTATAAAG GTAATTAACC ACTTAATCTA -6654
 AATAAGAAC AATGTATGTG GGGTTTCTAG CTTCTGAAGA AGTAAAGTT ATGGCCAGCA TGGCAGAAAT -6584

Fig. 10

GTGAGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTTAACC -6514
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGGTGAAAGG -6444
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA -6374
 CACCGTCCTC TCATTACCGG TGCTTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TTCGTGTTTG -6304
 GTTAAACTTA ATCTGTATGA ATCTGAAAC GAAAAATGGT GGTGATTTCC TCCAGAAGAA TTAGAGTACC -6234
 TGGCAGGAAG CAGGTGGCTC TGTGGACCTG AGCCACTTCA ATCTTCAAGG GTCTCTGGCC AAGACCCAGG -6164
 TGCAAGGCAG AGGCCTGATG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT -6094
 CGTTGGTGAG CAGCGCATGA AGTGCCCTTA TTTACGCTTT GCAAAAGATTG CTCTGGATAC CATCTGGAAG -6024
 AGGCGGCCAG CGGGAATGCA AGGAGTCAGA AGCCTCTGTC TCAAACCCAG GCCAGCAGCT ATGGCCGCCA -5954
 CCCGGGCGTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTCACCTGA -5884
 AGCAGTGACC AAGGTGTATT CTGAGGGAAG CTTGAGTTAG GTGCCTTCTT TAAACAGAA AGTCATGGAA -5814
 GCACCTTCT CAAGGGAAAA CCAGACGCCC GCTCTGCGGT CATTTACCTC TTTCTCTCT CCCTCTCTTG -5744
 CCCTCGCGGT TTCTGATCGG GACAGAGTGA CCCCGTGGA GCTTCTCCGA GCCCGTGCTG AGGACCCCTC -5674
 TGCAAGGGG TCCACAGACC CCCGCCCTGG AGAGAGGAGT CTGAGCCTGG CTTAATAACA AACTGGGATG -5604
 TGGCTGGGG CGGACAGCGA CGGCGGGATT CAAAGACTTA ATTCATGAG TAAATCAAC CTTTCCACAT -5534
 CCGAATGGAT TTGGATTTTA TCTTAATATT TCTTAAAT TCATCAAATA ACATTGAGG ATGCAGAAAT -5464
 CCAAAGGCGT AAAACAGGAA CTGAGCTATG TTTGCCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394
 ATTTTTCGCC CTAAGTACTT TTTATTGGTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAA GTACACGAGG -5324
 AGAGGCTGG GCGGCAGGGC TATGAGCAGC GCAGGGCCAC CGGGGAGAGA GTCCCCGGCC TGGGAGGCTG -5254
 ACAGCAGGAC CACTGACCGT CCTCCCTGGG AGCTGCCACA TTGGGCAACG CGAAGGCGGC CACGCTGCGT -5184
 GTGACTCAGG ACCCATACC GGCTTCTGCG GCCACCCAC ACTAACCCAG GAAGTCACGG AGCTCTGAAC -5114
 CCGTGGAAC GAACATGACC CTTGCCTGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA GTGGTGTGCA -5044
 GGAAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATTATT CATCTTCACC -4974
 CCAAGGACT GAATGATTCC AGCAACTTCT TCGGTGTGA CAAGCCATGA CAAACTCAG TACAACACC -4904
 ACTCTTTTAC TAGGCCACA GAGCAGGSC CACACCCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT -4834
 GCTTTCAGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTCC TCTAGACTAG TAGACCCTGG -4764
 CAGGCACTCC CCCAGATTCT AGGGCTGGT TGTGCTTCC CGAGGGCGCC ATCTGCCCTG GAGACTCAGC -4694
 CTGGGTGCC AACTGAGGC CAGCCCTGTC TCCACACCCT CGCCTCCAG GCCTCAGCTT CTCCAGCAGC -4624
 TTCTAAACC CTGGGTGGGC CGTGTTCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT TGTCTCAGCG -4554
 ACGTCTGTC CACGTTTCTT CCTCACATGG GGTGTCTGTC TCCTTCCCCA ACACCTCACAT GGTGTAAGG -4484
 GAGGAGATTC TGCGCCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC GATGCAGGTT -4414
 CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCGCTCTC CTGTCTCTG CCGGGGCGCT -4344
 CCGGTGTGTT CTTCTGTTTC TGTGCTCCTT TCCACGTCCA GCTGCGTGTG TCTCTGCCCG CTAGGGTCTC -4274
 GGGGTTTTTA TAGGCATAGG ACGGGGCGT GGTGGGCCAG GGCCTCTTG GGAATGCAA CATTTGGGTG -4204
 TGAAAGTAGG AGTGCTGTC CTCACCTAGG TCCACGGGCA CAGGCCTGGG GATGGAGCCC CCGCCAGGGA -4134
 CCGGCCCTTC TCTGCCAGC ACTTCTCTGC CCCCCTCCCT CTGGAACACA GAGTGGCAGT TTCCACAAGC -4064
 ACTAAGCATC CTCTTCCAA AAGACCCAGC ATTGGCACCC CTGGACATTT GCCCCACAGC CCGGGAATT -3994

c-Myc
 CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCGG CTGTTTTATT TTAATAGCTA -3924
 CAAAGCAGGG AAATCCCTGC TAAATGTCC TTTAAACAAAC TGGTTAAACA AACGGGTCCA TCCGCACGGT -3854
 GGACAGTTCC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG GAATTACGCT -3784
 GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAG AAAGAATTTC ACCCCATGGC -3714
 AGGGGAGTGG TTAGGGGGGT TAAGGACGGT GGGGGCGGCA GCTGGGGGCT ACTGCACAGA CCTTTTACTA -3644
 AAGCCAGTTT CCTGGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG TGGGGATGGG -3574
 GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CCTGGGCAGG ATAATGCTCT AGAGATGCCC -3504
 ACGTCTGAT TCCCCAAC CTGTGGACAG AACCOCGCCG GCCCAGGGC CTTTGCAGGT GTGATCTCCG -3434
 TGAGGACCCT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGCCCGAAAA GTAATCCAGG GGTTCGCGGA -3364
 AGAGGCGGGC AGGAGGGTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA GGCTGAAAAG -3294
 GAGGGGAGGG CCTCGAGCCC AGGCCTGCAA CGCCTCCAG AAGCTGGAAA AAGCGGGGAA GGGACCTCC -3224
 ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCTTAGCCCC ACCAGGGGCC ATCGTGGACC TCCGGCCTCC -3154
 GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGGG ATTTGCAGAA GCAACAGGAA -3084
 ACCCATGCAC TGTGAATCTA GGATTATTTT AAAACAAAGG TTTACAGAAA CATCCAAGGA CAGGGGTGAA -3014
 GTGCCCTCGG GCAAGGGCAG GGCAGGCACG AGTGATTTTA TTTAGCTATT TTATTTTATT TACTTACTTT -2944
 CTGAGACAGA GTTATGCTCT TGTGCCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC -2874
 CGTCTCCTGG GTTCAAGCAA TTCTCGTGCC TCAGCCTCCC AAGTAGCTGG GATTTACAGC GTGCACCACC -2804
 ACACCCGGCT AATTTTGTAT TTTAGTAGA GATGGGCTTT CACCATGTTG GTCAAGCTGA TCTCAAAATC -2734
 CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664
 GGCTTATTTA ACCATTTTAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT CATGGAGTTC -2594
 AATTTCCCTT TTAATCAGGA GTTACCCTCC TTTGATATT TCTGTAATTC TTCGTAGACT GGGGATACAC -2524
 CGTCTCTTGA CATATTCACA GTTTCTGTGA CCACCTGTTA TCCCATGGGA CCCACTGCGG GGGCAGCTGG -2454
 GAGGCTGCAG GCTTCAGGTC CCAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA ATCAGGGGCG -2384
 AAGTGTGGAC ACTGTCCTGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAAATA AAGTCCATCC -2314
 CTCTACTCT ACTGGGATTG AGCCCTTCC CTATCCCCCC CCAGGGGCAG AGGAGTTTCT CTCACTCTG -2244
 TGGAGGAAGG AATGATACTT TGTATTTTTT CACTGCTGGT ACTGAATCCA CTGTTTCATT TGTGTTGTTG -2174
 TTTGTTTTGT TTTGAGAGGC GGTTCCTACT TTGTTGCTCA GGCTGGAGGG AGTGCAATGG CGCGATCTTG -2104
 GCTTACTGCA GCCTCTGCCT CCCAGGTTCA AGTGATTCTC CTGCTTCCGC CTCCCATTG CTTGGGATTA -2034
 CAGGCACCCG CCACCATGCC CAGCTAATTT TTTGTATTTT TAGTAGAGAC GGGGGTGGGT GGGGTTCCAC -1964

ATGTTGGCCA GGCTGGTCTC GAACTTCTGA CCTACCTGCC TCGCCTCCTT AAAGTGGCTGG -1894
GATTACAGGT GTGAGCCACC ATGCCCAGCT CAGAATTAC TCTGTTTAGA AACATCTGGG TCTGAGGTAG -1824

CAAT-Box

GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTAAGCCAA TGATAGAATT TTTTATTGT TGTTAGAACA -1754

CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCAG CTTTTCAAAG ACACACTAAC TGCACCCATA -1684

ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTTCATTGAA TGCCGGGAGG CGTTTCCTCG CCATGCACAT -1614

GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTCT TCTCTCCCT CTTTTAAAT TGTGTTTTCT -1544

ATGTTGGCTT CTCTGCAGAG AACCAGTGTA AGCTACAAC TAACTTTTGT TGAACAAAT TTTCCAACC -1474

Spl
GCCCCTTTGC CCTAGTGGCA GAGACAATTC ACAAAACAG CCCTTTAAAA AGGCTTAGGG ATCCTAAGG -1404

GGATTCTAG AAGAGCGACC TGTAATCCTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334

GCCCAGGGAG GGTGCGAGGC CTGTTCAAAT GCTAGCTCCA TAAATAAAGC AATTTCCTCC GGCAGTTTCT -1264

GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTGTTAGC ATTTCAAGTGT TTGCCGACCT CAGCTACAGC -1194

ATCCCTGCAA GGCTCGGGA GAGCCCAAG TTTCTCGCCC CCTTAGATCC AAACCTTGAGC AACCCGGAGT -1124

CTGGATTCTT GGGAAGTCCT CAGCTGTCCT GCGGTTGTGC CGGGGCCCA GGTCTGGAGG GGACCAAGTG -1054

CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CGGGCCTCCT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT -984

GCTTGAGACC CGAGGCTGCC CTCACCCCTG TCGGGGCGGG ATGTGACCAG ATGTTGGCCT CATCTGCCAG -914

ACAGAGTGCC GGGGCCCAGG GTCAAGGCCG TTGTGGCTGG TGTGAGGCGC CCGGTGCGCG GCCAGCAGGA -844

CCAC-Box **Spl**

GCGCCTGGCT CCATTCCCA CCCTTCTCG ACGGGACCGC CCCGGTGGGT GATTAACAGA TTTGGGGTGG -774

TTTGCTCATG GTGGGGACCC CTCGCCGCT GAGAACCTGC AAAGAGAAAT GACGGGCTTG TGTCAGGAG -704

CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCGCT GCGCGTCCAG GGAGCAATGC -634

AP-2

GTCTCGGGT TCGTCCCCG CCGCTCTAC GCGCCTCGT CCTCCCTTC ACGTCCGGCA TTCGTGGTGC -564

CCGGAGCCCG ACGCCCCGCG TCCGGACCTG GAGGCAGCCC TGGGTCTCCG GATCAGGCCA GCGGCCAAAG -494

GGTCGCCGCA CGCACCTGT CCCAGGGCCT CCACATCATG GCCCCCTCCT CGGGTTACCC CACAGCCTAG -424

GCCGATTCTG CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCT GGGAGCGCGA GCGGGCGCG **Spl** -354

GCGCGGGAAG CGCGGCCAG ACCCCCCGGT CGGCCGGAG CAGCTGCGCT GTCGGGGCCA GGCCGGGCTC -284

CCAGTGGATT CGCGGGCACA GACGCCCAGG ACCGCGCTCC CACGTGCGG GAGGGACTGG GGACCCGGGC -214

ACCGTCTCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCC GCCCCTGCC GACCCCTCCC -144

GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCC CCCCCTTCCTT TCCGCGGCC CGCCCTCTCC **Spl** -74

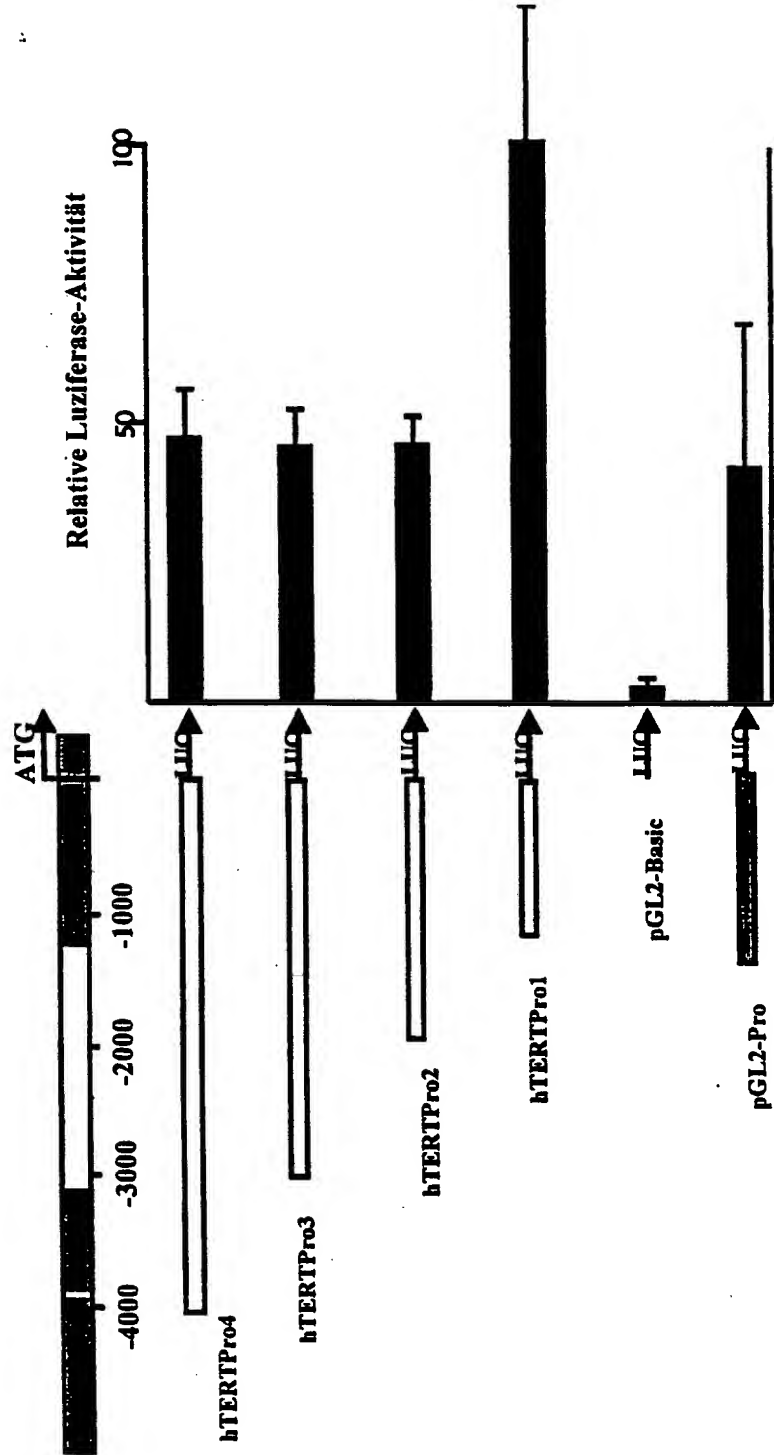
c-Myc

TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCG CACGTGGGAAG CCCTGGCCCC GGCCACCCCC -4

c-Myc

GCGATG

Fig.: 11



1 / 18

SEQUENZPROTOKOLL

<110> Bayer AG

5 <120> Regulatorische DNA-Sequenzen aus der 5i-Region vom Gen
der humanen katalytischen Telomerase-Untereinheit und
deren diagnostische und therapeutische Verwendung

10 <130> LeA32805-Ausland

<140>
<141>

15 <160> 20

<170> PatentIn Vers. 2.0

<210> 1
<211> 5126
20 <212> DNA
<213> Homo sapiens

<400> 1

25 gagctctgaa ccgtggaaac gaacatgacc cttgcctgcc tgcttccctg ggtgggtcaa 60
gggtaataaa gtgggtgtgca ggaaatggcc atgtaaatta cagcactctg ctgatgggga 120
ccgttccttc catcattatt catcttcacc cccaaggact gaatgattcc agcaacttct 180
tcgggtgtga caagccatga caaaactcag tacaacacacc actcttttac taggcccaca 240
gagcacgggc cacaccctcg atataatga agtccaggag agatgaggct gctttcagcc 300
accaggctgg ggtgacaaca gcggctgaac agtctgttcc tctagactag tagaccctgg 360
30 caggcactcc cccaaattct agggcctggg tgctgcttcc cgaggggcgc atctgcctcg 420
gagactcagc ctgggggtgcc aactgaggc cagccctgtc tccacaccct ccgcctccag 480
gcctcagctt ctccagcagc ttcctaaacc ctgggtgggc cgtgttccag cgctactgtc 540
tcacctgtcc cactgtgtct tgtctcagcg acgtagctcg caccgttctc cctcacatgg 600
ggtgtctgtc tcttcccca aactcacat gcgttgaagg gaggagattc tgcgcctccc 660
35 agactggctc ctctgagcct gaacctggct cgtggccccc gatgcaggtt cctggcgctc 720
ggctgcacgc tgacctccat ttccaggcgc tcccgtctc ctgtcatctg ccggggcctg 780
ccgggtgtgt cttctgttcc tgtgtctctt tccacgtcca cgtcgtgtgt tctctgccc 840
ctagggtctc ggggttttta taggcatagg acgggggcgt ggtggggcag ggcgctcttg 900
40 ggaatgcaa catttgggtg tgaagttagg agtgctgtc ctacactagg tccacgggca 960
caggcctggg gatggagccc ccgcccttc tctgccagc actttcctgc 1020
ccccctccct ctggaacaca gattggcagt ttccacaagc actaagcacc ctcttcccaa 1080
aagaccagc attggcacc ctggacattt gccccacagc cctgggaatt cactgtgacta 1140
cgacacatcat gtacacactc ccgtccacga ccgacccccg ctgttttatt ttaatagcta 1200
caaagcaggg aaatccctgc taaaatgtcc ttaacaaac tggttaaaca aacgggtcca 1260
45 tccgcacggg ggacagtcc tcacagtga gaggaaatg ccgtttataa agcctgcagg 1320
catctcaagg gaattacgct gattcaaaac tgccacctcc atgggatacg tacgcaacat 1380
gctcaaaaag aaagaatttc accccatggc aggggagtg ttaggggggt taaggacggt 1440
ggggggcgga gctgggggct actgcacgca ccttttacta aagccagttt cctggttctg 1500
atggtattgg ctcatgtatg ggagactaac cataggggag tggggatggg ggaacccgga 1560
50 ggctgtgcca tctttgccat gcccgagtgt cctgggcagg ataatgctct agagatgcc 1620
acgtcctgat tcccccaaac ctgtggacag aacccgcccg gcccaggggc ctttgagggt 1680
gtgatctccg tgaggacctt gagggtctgg atccttcggg actacctgca ggcccgaaaa 1740
gtaatccagg ggttctggga agaggcgggc agggagggtca gaggggggca gcctcaggac 1800
gatggaggca gtcagtctga ggctgaaaag ggagggaggg cctcgagccc aggcctgcaa 1860
55 gcgcctccag aagctggaaa aagcggggaa gggaccctcc acggagcctg cagcaggaa 1920
gcacggctgg cccttagccc accaggggcc atcgtggacc tccggcctcc gtgcatagg 1980
agggcactcg cgtgcctctt ctacatgaa gtgtgtgggg atttgcagaa gcaacaggaa 2040
acccatgcac tgtgaatcta ggattatttc aaaacaaagg ttacagaaa catccaagga 2100
cagggtgaa gtgcctccgg gcaagggcag ggcaggcacg agtgatttta tttagctatt 2160
60 ttattttatt tactacttt ctgagacaga gttatgctt tgtgtcccag gctggagtgc 2220
agcgcctga tcttggctca ctgcaacctc cgtctcctgg gttcaagcaa ttctcgtgcc 2280
tcagcctccc aagtagctgg gatttcaggc gtgcaccacc acaccggcct aattttgtat 2340
ttttagtaga gatgggcttt caccatgttg gtcaagctga tctcaaaatc ctgacctcag 2400
gtgatccgcc cactcagcc tcccaagtgt ctgggattac aggcattgag cactgcacct 2460

2 / 18

5 ggccatttta accattttta aacttccctg ggctcaagtc acacccactg gtaaggagtt 2520
 catggagttc aatttcccct ttactcagga gttaccctcc ttgtatattt tctgtaattc 2580
 ttctgtagact ggggatacac cgtctcttga catattcaca gtttctgtga ccacctgtta 2640
 tcccatggga cccactgcag gggcagctgg gaggtctgag gcttcaggtc ccagtggggt 2700
 tgccatctgc cagtagaaac ctgatgtaga atcagggcgc aagtggtggac actgtccctga 2760
 atctcaatgt ctcatgtgtg gctgaaacat gtgaaatta aagtcacatcc ctctactctc 2820
 actgggattg agccccctcc ctatccccc ctaggggagc aggagttcct ctactcctg 2880
 tggagggaagg aatgatactt tgttattttt cactgctggt actgaatcca ctgtttcatt 2940
 10 tgttggtttg tttgttttgt tttgagaggg ggtttcactc ttgttgctca ggctggaggg 3000
 agtgcaatgg cgcgactctg gcttactgca gcctctgcct cccagggtca agtgattctc 3060
 ctgcttccgc ctcccatttg gctgggatta caggcaccgc ccaccatgcc cagctaattt 3120
 tttgtatttt tagtagagac gggggtgggt ggggttcacc atgttggtcca ggctgggtcc 3180
 gaacttctga cctcagatga tccactgcc tctgcctcct aaagtgtgtg gattacaggt 3240
 gtgagccacc atgcccagct cagaatttac tctgtttaga aacatctggg tctgaggtag 3300
 15 gaagctcacc ccactcaagt gttgtgtgtg ttttaagcaa tgatagaatt tttttattgt 3360
 tgttagaaca ctcttgatgt tttacactgt gatgactaag acatcatcag cttttcaaag 3420
 acacactaac tgcaccata atactgggt gtcttctggg tatcagcaat ctctattgaa 3480
 tgccgggagg cgtttctctg ccatgcacat ggtgttaatt actccagcat aatcttctgc 3540
 20 ttcatttctc tctcttccct cttttaaaat tgtgttttct atgttggtct ctctgcagag 3600
 aaccagtgtg agctacaact taacttttgt tggaaacaaat ttccaaacc gccctttgct 3660
 cctagtggca gagacaattc acaaacacag ccttttaaaa aggcttaggg atcactaagg 3720
 ggatttctag aagagcgacc tgaatccta agtatttaca agacgaggt aacctccagc 3780
 gagcgtgaca gccagggag ggtgcgaggc ctgttcaaat gctagctcca taaataaagc 3840
 25 aatttctctc ggcagtttct gaaagtagga aaggttacat ttaagggtgc gtttcttagc 3900
 atttcagtgt ttgccgacct cagctacagc atccctgcaa ggctctggga gaccagaag 3960
 tttctcgcgc ccttagatcc aaacttgagc aacctggagt ctggattcct gggaagtcct 4020
 cagctgtcct gcggttgtgc cggggcccca ggtctggagg ggaccagtgg ccgtgtggct 4080
 tctactgctg ggctggaggt cgggctcct agctctgcag tccgagggtt ggagccaggt 4140
 30 gcctggacc ctaggctgcc ctccaccctg tgcgggcggg atgtgaccag atgttggtct 4200
 catctgccag acagagtgcc ggggcccagg gtcaaggccg ttgtggctgg tgtgaggcgc 4260
 ccggtgcgag gccagcagga gcgcctggct ccatttccca ccttctctcg acgggaccgc 4320
 cccggtgggt gattaacaga tttggggtgg tttgctcatg gtggggacc ctcgcccct 4380
 35 gagaacctgc aaagagaaat gacgggcctg tgtcaaggag ccaagtcgc ggggaagtgt 4440
 tgtagggagg cactccggga ggtcccgcgt gccctccag ggagcaatgc gtccctcgggt 4500
 tctccccag ccgctgtac gcgcctccgt cctcccttc acgtccggca tctgtgtgct 4560
 ccggagccc cgcgccgcg tccggacctg gaggcagccc tgggtctccg gatcaggcca 4620
 gcggccaaag ggtgcgcgca cgcacctgtt cccaggccct ccacatcatg gcccctccct 4680
 cggtgtaccc cacagcctag gccgattcga cctctctccg ctggggccct cgtgtggctc 4740
 40 cctgcaccct gggagcgcga gcggcgcgcg ggccgggaaag gcgggcccag accccgggt 4800
 ccgcccggag cagctgcgct gtccggggcca ggccgggctc ccagtggatt cgcgggcaca 4860
 gacgcccagg accgcgctcc ccacgtggcg gagggactgg ggaccgggc acccgtcctg 4920
 ccccttcacc ttccagctcc gccctctccg cgcggacccc gccccgtccc gaccctccc 4980
 45 gggctcccgc cccagccccc tccgggcccet cccagcccet ccccttccct tccgcccgc 5040
 cgccctctcc tgcggcgcg agtttcaggc agcgtctgct cctgtctgct acgtgggaag 5100
 ccctggcccc ggccaccccc gcgatg 5126

<210> 2

<211> 4042

<212> DNA

50 <213> Homo sapiens

<400> 2

55 gtttcaggca gcgctgcgtc ctgctgcgca cgtgggaagc cctggccccg gccacccccg 60
 cgatgccgag cgctccccgc tgcgagagcg tgcgctccct gctgcgcagc cactaccgag 120
 aggtgctgcc gctggccacg ttctgtgcggc gcctggggcc ccagggtctg cggctggtgc 180
 agcgcgggga cccggcggtt ttccgcgcgc tgggtggcca gtgcctggtg tgcgtgccct 240
 gggacgcacg gccgcccccc gccgccccct ccttccgcca ggtgtcctgc ctgaaggagc 300
 tgggtggccc agtgctgcag aggtgtgtcg agcgcggcgc gaagaacgtg ctggccttcg 360
 60 gcttcgcgct gctggacggg gcccgcgggg gccccccga ggcccttacc accagcgtgc 420
 gcagctacct gcccaacacg gtgaccgacg cactgcgggg gagcggggcg tgggggctgc 480
 tgctgcgccc cgtgggagac gacgtgctgg ttacactgct ggacacgtgc gcgtctcttg 540
 tgctggtggc tcccagctgc gcctaccagg tgtgcgggccc gccgctgtac cagctcggcg 600
 ctgccactca ggcccggccc ccgccacacg ctagtggacc ccgaaggcgt ctgggatgag 660
 65 aacgggccc gaaacatagc gtccaggagg ccgggtccc cctgggccc ctagcccccg 720
 gtgcgaggag gcgcgggggc agtgccagcc gaagtctgct gttgcccaag aggccaggc 780

3 / 18

gtggcgctgc ccctgagccg gagcggagcg ccgttgggca ggggtccctgg gccaccccg 840
 gcaggacgcg tggaccgagt gaccgtggtt tctgtgtggt gtcacctgcc agaccgcgcg 900
 aagaagccac ctctttggag ggtgctctct ctggcagcg ccactccac ccatccgtgg 960
 gccgccagca ccacgcgggc ccccatcca catcgcgcc accacgtccc tgggacacgc 1020
 5 cttgtccccc ggtgtacgcc gagaccaagc acttctctta ctctcaggg gacaaggagc 1080
 agctgcggcc ctctcttcta ctacgtcttc tgaggccag cctgactggc gctcggaggc 1140
 tcgtggagac catctttctg ggttccaggg cctggatgcc agggactccc cgcagggttg 1200
 ccgcctgcc ccacgcgtac tggcaaatgc ggcctctgtt tctggagctg cttgggaacc 1260
 acgcgcagtg cccctacggg gtgctctcca agacgcactg cccgctgcga gctgcggtca 1320
 10 ccccgagcagc cgggtgtctgt gcccgggaga agccccaggg ctctgtggcg gcccccaggg 1380
 aggaggacac agacccccgt cgctgtgtgc agctgtctcg ccagcacagc agccccggc 1440
 aggtgtacgg cttcgtgcgg gcctgcctgc gccggctggg gcccccaggc ctctggggct 1500
 ccaggcaca cgaacggcgc ttctcagga acaccaagaa gttcatctcc ctgggggaagc 1560
 atgccaagct ctctgtcgag gagctgacgt ggaagatgag cgtgcgggac tccgtctggc 1620
 15 tgccgagggg cccaggggtt ggctgtgttc cggccgcaga gcacctctg cgtgaggaga 1680
 tcctggccaa gttcctgcac tggctgatga gtgtgtacgt cgtcgagctg ctcagggtctt 1740
 tcttttatgt cacggagacc acgtttcaaa agaacaggct ctttttctac cgggaagagt 1800
 tctggagcaa gttgcaaaag attggaatca gacagcactt gaagagggtg cagctcggg 1860
 agctgtcgga agcagaggtc aggcagcatc gggaagccag gcccgccctg ctgacgtcca 1920
 20 gactccgctt catccccag cctgacgggc tgcggccgat tgtgaacatg gactacgtcg 1980
 tgggagccag aacgttcccg agagaaaaga gggccgagcg tctcacctcg aggggtgaag 2040
 cactgttcag cgtgtcaac tacgagcggg cgcggcgccc cggcctctg ggcgctctg 2100
 tgctgggctt ggacgatac caccgggctt ggcgcacctt cgtgtgctg gtgcgggccc 2160
 aggacccgcc gcctgagctg tactttgtca aggtggatgt gacggcgcg tacgacacca 2220
 25 tccccagga caggctcacg gaggtcatcg ccagcatcat caaacccag aacacgtact 2280
 gcgtgcgtcg gtatgccgtg gtccagaagg cgcgccatgg gcacgtccg aaggcctca 2340
 agagccacgt ctctacctt acagacctc agccgtacat gcgacagttc gtgggtcacc 2400
 tgccaggagac cagcccgctg agggatgccc tctcatcga gcagagctcc tccctgaatg 2460
 agggcagcag tggcctcttc gacgtcttcc tacgttctat gtgccaccac gccgtgcgca 2520
 30 tcaggggcaa gtcctacgtc cagtgcagg ggaatccgca gggctccatc ctctccacgc 2580
 tgctctcgag cctgtgtac ggcgacatgg agaacaagct gtttgcgggg attcggcggg 2640
 acgggtgctt cctgctgtt gtggatgatt tcttgttgg gacacctcac ctacccacg 2700
 cgaaaacctt cctcaggacc ctgggtccgag gtgtccctga gtatggctgc gtggtgaact 2760
 tgccgaagac agtggtaaac ttccctgtag aagacgaggc cctgggtggc acggcttttg 2820
 35 ttcagatgcc ggccacggc ctattccctt ggtgcggcct gctgctggat acccggaacc 2880
 tggaggtgca gagcgactac tccagctatg cccggacctc catcagagcc agtctcact 2940
 tcaaccggcg cttcaaggct gggaggaaca tgcgtcgcaa actctttggg gtcttgcggc 3000
 tgaagtgtca cagcctgtt ctggatttgc aggtgaacag cctccagacg gtgtgcacca 3060
 40 acatctacaa gatcctcctg ctgcaggcgt acagggttca cgcagtgtg ctgcagctcc 3120
 catttcatca gcaagtttgg aagaacccca catttttctt gcgcgtcatc tctgacacgg 3180
 cctcctctg ctactccatc ctgaaagcca agaacgcagg gatgtcgctg ggggccaagg 3240
 gcgcgcggcg cctctgtccc tccgagggcg tgcaagtggt gtgccacca gcaattctgc 3300
 tcaagctgac tcgacaccgt gtcacctacg tgccactcct ggggtcactc aggacagccc 3360
 45 agacgcagct gactcggaag ctcccgggga cgcagctgac tgccctggag gccgcagcca 3420
 acccggcact gccctcagac ttcaagacca tcttgactg atggccaccc gccacagcc 3480
 agggcgagag cagacaccag cagccctgtc acgcggggt ctacgtcca gggaggagg 3540
 ggcggccac acccaggccc gcaccgctgg gactctgagg cctgagttag tgtttggccg 3600
 aggcctgcat gtccggctga aggtgagtg tccggctgag gcctgagcga gtgtccagcc 3660
 50 aagggctgag tgtccagcac acctgcccgtc ttcaacttccc cacaggctgg cgtctcggtc 3720
 caccacaggg ccagcttttc ctaccagga gcccggttcc cactccccac ataggaatag 3780
 tccatcccca gattcgccat tgttcaaccc tcgcccctgcc ctcttttgcc tccaccccc 3840
 accatccagg tggagacctt gagaaggacc ctgggagctc tgggaatttg gactgaccaa 3900
 aggtgtgccc tgtacacagg cgaggacctt gcacctggat ggggggtccc gtgggtcaaa 3960
 55 ttggggggag gtgctgtggg agtaaaatc tgaatatatg agtttttcag ttttgaaaaa 4020
 aaaaaaaaa aaaaaaaaa aa 4042

<210> 3
 <211> 11276
 <212> DNA
 60 <213> Homo sapiens

<400> 3
 acttgagccc aagagttcaa ggctacgggt agccatgatt gcaacaccac acgccagcct 60
 tgggtgacaga atgagacctt gtctcaaaaa aaaaaaaaaa aattgaaata atataagca 120
 65 tcttctcttg ccacagtga acaaaaccag aatcaacaa caagaggaat tttgaaaact 180

atacaaacac atgaaaaatta aacaatatatc ttctgaatga ccagtgcagtc aatgaagaaa 240
 tcaaaaagga aattgaaaaa ttattttaag caaatgataa cggaaacata acctctcaaa 300
 acccacggta tacagcaaaa gcagtgcata gaaggaagtt tatagctata agcagctaca 360
 tcaaaaaagt agaaaaagcca ggcgcagtggt ctcagtcctg taatcccagc actttgggag 420
 5 gccaaaggcgg gcagatcgcc tgaggtcagg agttcgagac cagcctgacc aacacagaga 480
 aaccttgctcg ctactaaaaa taaaaaatta gctgggcatg gtggcagatg cctgtaatcc 540
 cagctactcg ggaggctgag gcaggataac cgcttgaacc caggaggtgg aggttgcggt 600
 gagccgggat tgcgccattg gactccagcc tgggtaacaa gagtgaacc ctgtctcaag 660
 aaaaaaaaaa aagtagaaaa acttaaaaat acaacctaat gatgcacctt aaagaactag 720
 10 aaaagcaaga gcaactaaa cctaaaattg gtaaaagaaa agaaataata aagatcagag 780
 cagaaataaa tgaaactgaa agataacaat acaaaagatc aacaaaatta aaagtgtggt 840
 ttttgaagaag ataaacaaaa ttgacaaacc ttgtcccaga ctaagaaaaa aggaagaag 900
 acctaaataa ataaagtccag agatgaaaaa agagacatta caactgatac cacagaat 960
 caaaggatca ctagggtgta ctatgagcaa ctgtacacta ataaattgaa aaacctagaa 1020
 15 aaaatagata aatttcctaga tgcatacaac ctaccaagat tgaacctga agaaatccaa 1080
 agcccaaaac gaccaataac aataatggga ttaaagccat aataaaaagt ctcttagcaa 1140
 agagaagccc aggacccaat ggcttccctg ctggatttta ccaatcattt aaagaagaat 1200
 gaattccaat cctactcaaa ctattctgaa aaatagagga aagaataactt ccaactct 1260
 20 tctacatggc cagtattacc ctgattccaa aaccagacaa aaacacatca aaaacaaaca 1320
 aacaaaaaaa cagaaagaaa gaaaactaca ggccaatctc cctgatgaat actgatacaa 1380
 aaatcctcaa caaaacacta gcaaaccaaa ttaaacacaa ccttcgaaag atcattcatt 1440
 gtgatcaagt gggatttatt ccagggatgg aaggatggtt caacatagc aaatcaatca 1500
 atgtgatata tcatcccaac aaaatgaagt acaaaaacta tatgattatt tcactttatg 1560
 cagaaaaagc atttgataaa attctgcacc ctctcatgata aaaaccttca aaaaaccagg 1620
 25 tatcaagaa acatacaggg caggcacagt ggcctcacacc tgcgatccca gcaactctggg 1680
 aggccaaggt gggatgattg cttgggcccga ggagtttgag actagcctgg gcaacaaaat 1740
 gagacctggt ctacaaaaaa cttttttaa aaatagcca ggcattgatgg catatgctgt 1800
 tagtcccagc tagtctggag gctgagggtg gagaatcact taagcctagg aggtcgaggc 1860
 30 tgcagtgagc catgaaactt tcaactgtact ccagcctaga caacagaaca agaccccaat 1920
 gaataagaag aaggagaagg agaagggaaga agggaggagg aaggaggagg gaggagaagg 1980
 agggagggtg gaagaagtgga aaggggaagg ggaagggaag gaggaagaag aagaacata 2040
 tttcaacata ataaaagccc tatatgacag accgaggtag tattatgagg aaaaactgaa 2100
 agccttttct ctaagatctg gaaaatgaca agggcccaat ttcaccactg tgattcaaca 2160
 tagtactaga agtcctagct agagcaatca gataagagaa agaaataaaa ggcatccaaa 2220
 35 ctggaaagga agaagtcaaa ttatctgtt tgcagatgat atgatcttat atctgaaaa 2280
 gacttaagac accactaaaa aactattaga gctgaaattt ggtacagcag gatacaaaa 2340
 caatgtacaa aaatcagtag tatttctata ttccaacagc aaacaatctg aaaaagaaac 2400
 caaaaaagca gctacaaata aaattaaaca gctagggaatt aaccaaagaa gtgaaagatc 2460
 40 tctacaatga aaactataaa atgttgataa aagaaattga agagggcaca aaaaagaaa 2520
 agatattcca tgttcataga ttggaagaat aaatactgtt aaatgtcca tactaccaca 2580
 agcaatttac aaattcaatg caatccctat taaaatacta atgacgttct tcacagaaat 2640
 agaagaaaca attctaagat ttgtacagaa ccacaaaaga cccagaatag ccaagctcat 2700
 cctgaccaa aagaacaaaa ctggaagcat cacattacct gacttcaaat tatactacaa 2760
 45 agctatagta acccaacta catggtactg gcataaaaac agatgagaca tggaccagag 2820
 gaacagaata gagaatccag aaacaaatcc atgcatctac agtgaactca tttttgacaa 2880
 aggtgccaa aacatacttt ggggaaaaga taatctcttc aataaatggt gctggaggaa 2940
 ctggatatcc atatgcaaaa taacaatact agaactctgt ctctcaccat atacaaaagc 3000
 aaatcaaaat ggatgaaagg cttaaatcta aaacctcaaa ctttgcaact actaaaagaa 3060
 50 aacaccggag aaactctcca ggacattgga gtgggcaaa acttcttgag taattccctg 3120
 caggcacagg caaccaaagc aaaaacagac aaatgggagc atatcaagtt aaaaagcttc 3180
 tgcccagcaa aggaaacaat caacaaagag aagagacaac ccacagaatg ggagaatata 3240
 tttgcaaaact attcatctaa caaggaatta ataaccagta tatataagga gctcaaaacta 3300
 ctctataaga aaaaacaccta atagctgat tttcaaaaat aagcaaaaaga tctgggtaga 3360
 55 catttctcaa aataagtcata acaaatggca aacaggcatc tgaatatgtg ctcaacacca 3420
 ctgatcatca gagaaatgca aatcaaaact actatgagag atcatctcat cccagttaaa 3480
 atggctttta ttcaaaagac aggcaataac aaatgccagt gaggtgtggt ataaaaggaa 3540
 acccttggac actgttgggt ggaatggaaa ttgctaccac tatggagaa agtttgaaag 3600
 ttcctcaaaa aactaaaaat aaagctacca tacagcaatc ccatgtctag gtatatactc 3660
 caaaaaaggg aatcagtgta tcaacaagct atctccactc ccacatttac tgcagcactg 3720
 60 ttcatagcag ccaagggttg gaagcaacct cagtgtccat caacagacga atggaaaaag 3780
 aaaaatgtgt gcacatacac aatggagtac tacgcagcca taaaaagaa tgagatcctg 3840
 tcagttgcaa cagcatgggg ggcactgggt agtatgttaa gtgaaataag ccaggcacag 3900
 aaagacaaac ttttctatgt ctcccttact tgtgggagca aaaaattaaa caattgacat 3960
 65 agaaatagag gagaatgggt gttctagagg ggtgggggac aggggtgacta gagtcaacaa 4020
 taattttatt tatgttttaa aataactaaa agagtataat tgggttgttt gtaacacaaa 4080

gaaaggataa atgcttgaag gtgacagata ccccatTTac cctgatgtga ttattacaca 4140
 ttgtatgcct gtatcaaaat atctcatgta tgctatagat ataaacccta ctatattaaa 4200
 aattaaaaat ttaatggcca ggcacggtgg ctcatgtccg taatcccgag accttgggag 4260
 5 gccgagggcg gtggatcacc tgaggtcagg agtttgaaac cagtctggcc accatgatga 4320
 aacctgtct ctactaaaga tacaaaaatt agccaggcgt ggtggcacat acctgtagtc 4380
 ccaactactc aggaggctga gacaggagaa ttgcttgaac ctgggaggcg gaggttgcag 4440
 tgagccgaga tcatgccact gcactgcagc ctgggtgaca gagcaagact ccatctcaaa 4500
 acaaaaaaaa aaaaaagaag attaaaaatt taatttttat gtaccgtata aatataact 4560
 ctactataatt agaagttaaa aattaaaaa attataaaag gtaattaaac acttaactta 4620
 10 aaataagaac aatgtatgtg ggggttctag cttctgaaga agtaaaagt atggccacga 4680
 tggcagaat gtgaggagg aacagtggaa gttactgttg ttagacgctc atactctctg 4740
 taagtgaact aattttaacc aaagacaggc tgggagaagt taaaggagca ttctataagc 4800
 cctaaaaaaa ctgctaataa tggtgaaagg taatctctat taattaccaa taattacaga 4860
 tatctctaaa atcgagctgc agaattggca cgtctgatca caccgtccct tcattcacgg 4920
 15 tgcttttttt cttgtgtgtg tggagatttt cgattgtgtg ttctgttttg gttaaaactta 4980
 atctgtatga atcctgaaac gaaaaatgg ggtgatttcc tccagaagaa ttagagtacc 5040
 tggcagggaag caggtggctc tgtggacctg agccacttca atcttcaagg gtctctggcc 5100
 aagaccagg tgcaaggcag aggcctgatg acccgaggac aggaaagctc ggaagggaag 5160
 gggcgatgag aagcctgcct cgttgggtgag cagcgcatga agtgccctta ttacgcttt 5220
 20 gcaagatttg ctctggatac catctggaaa agcgcgccag cgggaatgca aggagtcaga 5280
 agcctctctg tcaaacccag gccagcagct atggcgccca cccggcgctg tgccagaggg 5340
 agaggagtca aggcacctcg aagtatggct taaatctttt ttccacctga agcagtgacc 5400
 aaggtgtatt ctgagggaag cttgagttag gtgccttctt taaaacagaa agtcatgga 5460
 gcacccttct caagggaaaa ccagacgccc gctctgcggt catttacctc ttctctctct 5520
 25 cctctcttg cctctcggtt tctgactcg gacagagtga ccccgctgga gcttctccga 5580
 gcccggtctg aggaacctt tgcaaggggc tccacagacc cccgcctctg agagaggagt 5640
 ctgagcctgg cttaataaca aactgggatg tggctggggg cggacagcga cggcgggatt 5700
 caaagactta attccatgag taaattcaac ctttccacat ccgaatggat ttgattttaa 5760
 30 tcttaataat ttcttaaat tcatcaata acattcagga ctgcagaaat ccaaggcgct 5820
 aaaacaggaa ctgagctatg ttgccaagg tccaaggact taataacat gttagaggg 5880
 atttttcgcc ctaagtactt tttattggtt ttcataagggt ggttaggggt gcaagggaag 5940
 gtacacgagg agaggcctg gcggcagggc tatgagcacg cagggggcac cggggagaga 6000
 gtccccggcc tgggaggctg acagcaggac cactgacctt cctccctggg agctgccaca 6060
 35 ttgggcaacg cgaaggcgcc cagctgcgt gtgactcagg accccatacc ggtctctctg 6120
 gcccaccac actaacccag gaagtcaagg agctctgaac ccgtggaaac gaacatgacc 6180
 cttgccctgc tgcctccctg ggtgggtcaa gggtaatgaa tgggtgtgca ggaatggcc 6240
 atgtaaatca cagactctg ctgatgggga cgttctcttc catcattatt catcttcacc 6300
 cccaaggact gaatgattcc agcaacttct tgggtgtgta caagccatga caaaactcag 6360
 40 tacaacaccc actcttttgc taggccaca gagcacggsc cacacctg atataatag 6420
 agtccaggag agatgaggct gctttcagcc accaggctgg ggtgacaaca gcggctgaac 6480
 agtctgttcc tctagactag tagacctgg caggcactcc cccagattct agggcctggt 6540
 tgctgcttcc cgaggggccc atctgcctg gagactcagc ctgggtgccc acactgaggc 6600
 cagccctgtc tccacacct ccgctccag gcctcagctt ctccagcagc ttccataaac 6660
 45 ctgggtgggc cgtgttccag cgctactgtc tcacctgtcc cactgtgtct tgtctcagc 6720
 acgtagctcg caggttccct cctcacatgg ggtgtctgtc tcttcccca acactcacat 6780
 gcgttgaggg gaggagattc tgcgctccc agactggctc ctctgagcct gaacctggct 6840
 cgtggccccc gatgcagggt cctggcgctc ggctgcacgc tgacctccat ttccaggcgc 6900
 tccccgtctc ctgtcatctg ccggggcctg ccggtgtgtt cttctgtttc tgtgtcctt 6960
 50 tccacgtcca gctgcgtgtg tctctgccc ctagggtctc ggggttttta taggcatagg 7020
 acggggggcg ggtggggcag ggcgctcttg ggaatgcaa catttgggtg tgaagttagg 7080
 agtgccctgtc ctacactagg tccacgggca caggcctggg gatggagccc ccgccaggga 7140
 cccgcccctc tctgcccagc actttctctg cccctccct ctggaacaca gagggtgagt 7200
 ttccacaagc actaagcatc ctcttcccaa aagaccagc attggcacc ctggacattt 7260
 55 gcccacagc cctgggaatt cacgtgacta cgcacatcat gtacacactc ccgtccacga 7320
 ccgacccccg ctgttttatt ttaatagcta caaagcaggg aaatccctgc taaaatgtcc 7380
 tttacaaaac tgggttaaca aacgggtcca tccgcacggg ggacagttcc tcacagtga 7440
 gaggaacatg ccgtttataa agcctgcagg catctcaagg gaattacgct gagtcaaaac 7500
 tgccacctcc atgggatacg tacgcaacat gctcaaaaag aaagaatttc acccatggc 7560
 aggggagtg ttaggggggt taaggacggt gggggcgcca gctgggggct actgcacgca 7620
 60 ccttttacta aagccagtgt cctggtctg atggtattg ctcagttatg ggagactaac 7680
 cataggggag tggggatggg ggaacccgga ggctgtgcca tctttgccat gcccgagtgt 7740
 cctgggcagg ataagtctc agagatgccc acgtcctgat tcccccaaac ctgtggacag 7800
 aaccgcccgc gcccagggc ctttgcaagg gtgatctccg tgaaggacct gaggctctgg 7860
 65 atcctcggg actacctga ggcgggaaa ggttctggga agaggcgggc 7920
 aggaggggtca gaggggggca gcctcaggac gatggaggca gtcagtctga ggtgaaaag 7980

6 / 18

5 ggagggaggg cctcgagccc aggcctgcaa gcgcctccag aagctggaaa aagcggggaa 8040
 gggaccctcc acggagcctg cagcaggaaag gcacgggctg cccttagccc accagggccc 8100
 atcgtggacc tccggcctcc gtgccatagg agggcactcg cgctgcctct ctgcatgaa 8160
 gtgtgtgggg atttgacagaa gcaacaggaa acccatgcac tgtgaatcta ggattatttc 8220
 10 aaaaacaaagg tttacagaaa catccaagga cagggtgaa gtgcctccgg gcaagggcag 8280
 ggcaggcacg agtgatttta tttagctatt ttattttatt tacttacttt ctgagacaga 8340
 gttatgctct tgttgcctcg gctggagtgc agcggcatga tcttggctca ctgcaacctc 8400
 cgctccctgg gttcaagcaa ttctcgtgcc tcagcctccc aagtagctgg gatttcaggc 8460
 gtgcaccacc acaccggct aattttgtat ttttagtaga gatgggcttt caccatgttg 8520
 gtcaagctga tctcaaaatc ctgacctcag gtgatccgcc caccrcagcc tcccaagtg 8580
 ctgggattac aggcattgagc cactgcacct ggctatttta accattttaa aacttccctg 8640
 ggctcaagtc acaccactg gtaaggagtt catggagttc aatttccctt ttactcagga 8700
 gttaccctcc tttgatattt tctgtaattc tctgtagact ggggatacac cgtctcttga 8760
 15 catattcaca gtttctgtga ccacctgtta tccatggga cccactgcag gggcagctgg 8820
 gaggctgacg gcttcaggtc ccagtggggg tgccatctgc cagtagaaac ctgatgtaga 8880
 atcagggcgc aagtgtggac actgtcctga atctcaatgt ctcaagtgtg gctgaaacat 8940
 gtgaaatata aagtcacatc ctccctactct actgggattg agccctctcc ctatccctcc 9000
 ccagggcag aggagttcct ctcaactcctg tggaggaaag aatgatactt tgttattttt 9060
 cactgctggt actgaatcca ctgtttcatt tgttggtttg tttgttttgt tttgagaggc 9120
 20 ggtttcactc ttgttgctca ggctggaggg agtgcaatgg cgcgatcttg gcttacttga 9180
 gcctctgctt cccaggttca agtgattctc ctgcttccgc ctccatttg gctgggatta 9240
 caggcacccg ccaccatgcc cagctaatct tttgtatttt tagtagagac ggggggtggg 9300
 ggggttcacc atgttggcca ggctggcttc gaacttctga cctcagatga tccacctgcc 9360
 tctgctcctt aaagtgtgtg gattacaggt gtgagccacc atgcccagct cagaatttac 9420
 25 tctgtttaga aacatctggg tctgaggtag gaagctcacc ccactcaagt gttgtggtgt 9480
 ttttaagcaa tgatagaatt tttttattgt tgttagaaca ctcttgatgt tttactactg 9540
 gatgactaag acatcatcag cttttcaaaag acacactaac tgcaccata atactgggg 9600
 gtctctctgg tatcagcaat ctctcattgaa tgccgggagg cggttctctg ccatgcacat 9660
 30 ggtgttaatt actccagcat aatcttctgc ttccatttct tctcttccct cttttaaaat 9720
 tgtgttttct atgttggctt ctctgcagag aaccagtgtg agctacaact taacttttgt 9780
 tggacaaat tttccaaacc gcccttttgc ctagtgga gagacaattc aaaaacacag 9840
 ccctttaaaa aggccttaggg atcactaagg ggatttctag aagagcgacc tgtaatctta 9900
 agtatttaca agacgaggct aacctccagc gagcgtgaca gccagggag ggtgcgaggc 9960
 35 ctgttcaaat gctagctcca taaataaagc aatttccctc ggcagtttct gaaagttaga 10020
 aaggttacat ttaagggtgc gtttgttagc atttcagtgt ttgcccagct cagctacagc 10080
 atccctgcaa ggcctcggga gaccagaag tttctcgcct ccttagatcc aaacttgagc 10140
 aaccggaggt ctggattcct gggaaagtcct cagctgtcct gcggttgtgc cggggcccca 10200
 ggtctggagg ggaccagtg cctgtgtgct tctactgtgt ggctggaagt cgggctcctc 10260
 40 agctctcgag tccgaggctt ggagccaggt gcctggaccc cgaggctgcc ctccacctgt 10320
 tgcgggcggg atgtgaccag atgttggcct catctgccag acagagtgc gggggccagg 10380
 gtcaaggccg ttgtggtggt tgtgaggcgc ccggtgcgcg gccagcagga gcgcctggct 10440
 ccatttccca ccctttctcg acgggaccgc ccgggtgggt gattaacaga tttgggggtg 10500
 tttgtcatg gtggggaccc ctgcgcgcct gagaacctgc aaagagaaat gacgggctctg 10560
 45 tgtcaaggag cccaagtgc ggggaagtgt tgcaggagg cactccggga ggtccgcgt 10620
 gcccgctcag ggagcaatgc gtcctcgggt tctgtcccg ccgcgtctac gcgcctcctg 10680
 cctcccttc acgtccggca ttctgtgtgc ccggagcccg acgccccgcg tccggacctg 10740
 gaggcagccc tgggtctccg gatcaggcca gcggccaaag ggtcgcgcga cgcacctgt 10800
 cccaggccct ccacatcatg gccctccct cgggttacc caccagctag gccgattcga 10860
 50 cctctctcct ctggggccct cgtggcgct cctgcacct gggagcgca ggcgcgcg 10920
 ggcggggaag cgcggccag accccgggt cgcggcgag cagctgcgt gtcggggcca 10980
 ggcggggtc ccagtggatt cgcgggcaca gacggccag accgcgctcc ccacgtggcg 11040
 gagggactgg ggaccgggc acccgtcctg ccccttacc tccagctcc gccctcctcg 11100
 cgcggacccc gccccgtccc gacccctccc ggggtcccg cccagcccc tccgggccc 11160
 55 cccagccct ccccttccct tccggggccc cgcctctcc tccggggcg agtttcaggc 11220
 agcgtgcgt cctgctgcgc acgtgggaag ccctggcccc ggccaccccc gcgatg 11276

<210> 4
 <211> 104
 <212> DNA
 60 <213> Homo sapiens

<400> 4
 gtgggctcc ccggggtcgg cgtccggctg ggggtgaggg cggccggggg gaaccagcga 60
 catgcgagga gcagcgagc cgactcaggg cgcttcccc gcag 104

65

[illegible]

ctctagtaat ctagtaattc tttttttaa ttgctcttag tactgccaca ctgggcttct 3600
 tttgattagt attttccctgc tgtgtctgtt ttctgccttt aatttatata tatatatata 3660
 tttttttttt ttttgagaca gagtcttggg ctgtcgccca ggggtgagtgc agtgggtgta 3720
 ccacaggta gtgtaacttt taccttctgg cctgagccgt cctctcacct cagcctcctg 3780
 5 agtagctgga actgcagaca cgcaccgcta cacctggcta atttttaa attttctgga 3840
 gacagggtct tgcgtgtgtg cccaggctgg tctcaaactc ttggactcaa gggatccatc 3900
 tacctcggct tcccaaagt ctgaattaca ggcattgagcc accatgtctg gcctaatatt 3960
 caacactttt atattcttat agtgtgggta tgtcctgtta acagcatgta ggtgaatttc 4020
 caatccagtc tgacagtcgt tgtttaactg gataacctga tttattttca ttttttctgc 4080
 10 actagagacc cgcctgggtgc actctgattc tccacttgcc tgttgcatgt cctcgttccc 4140
 ttgtttctca ccacctcttg ggttgccatg tgcgtttcct gccgagtgtg tgttgatcct 4200
 ctctgtgctt cctgggtcact gggcatttgc ttttatttct ctttcttag tgttaccctc 4260
 tgatcttttt attgtcgttg tttgctttt tttattgaga cagtctcact ctgtcaccca 4320
 ggtctggagt taatggcaca atctcggtc actgcaacct ctgcctcctc ggttcaagca 4380
 15 gttctcattc ctcaacctga tgagttagct ggattacagg cgccaccac cagcctcctg 4440
 taatttttgt atttttagta gagataggct ttcaccatgt tggccaggct ggtctcaaac 4500
 tcttgacctc aagtgatctg cccgccttgg cctccacacg tgctgggatt acagggtgca 4560
 gccaccgtgc ccgcataacc ttgatctttt aaaatgaagt ctgaaacatt gctacccttg 4620
 tcttgagcaa taagaccctt agtgtatttt agctctggcc acccccagc ctgtgtgtctg 4680
 20 ttttccctgc tgacttagtt ctatctcagg catcttgaca ccccacaaag ctaagcatca 4740
 ttaatatgtt tttccgtgtt gagtgtttct gtactttgc ccccgccctg ctttctctcc 4800
 tttgttcccc gtctgtcttc tgtctcagge ccgcctctg ggggtccctt ccttctctt 4860
 tgcgtggttc ttctgtcttg ttattgctgg taaacccag ctttacctgt gctggcctcc 4920
 atggcatcta ggcagctccg gggacctctg cttatgatgc acagatgaag atgtggagac 4980
 25 tcacgaggag ggcggtcatc ttggcccggt agtgtctgga gcaccacgtg gccagcgttc 5040
 cttagccagt gagtgcagc aacgtccgct cggcctgggt tcagcctgga aaacccagg 5100
 catgtcgggg tctggtggct ccgcggtgtc gagtgtgaaa tcgcgcaaac ctgctgggtg 5160
 gcgcagctc tgacggtgct gctcggcggg ggagtgtctg cttcctcctt tctgtctggg 5220
 aaccaggaca aaggatgagg ctccgagccg ttgtcgccta acaggagcat gactgtgacc 5280
 30 atgtggataa ttttaaaatt tctaggctgg gcgcggtggc tcacgcctgt aatcccagca 5340
 ctttgggagg ccaaggcggg tggatcacga ggtcaggagg tcgagaccat cctggccaac 5400
 atgatgaac cccatctgta ctaaaaacac aaaaattagc tggcggtggg ggcgggtgct 5460
 tgtaatccca gctactcggg aggcctgaggc aggagaattg cttgaacctg ggagtggaa 5520
 gttgcagtg ggcagacctg caccactgca ctccagcctg gcaacacagc gagactctgt 5580
 35 ctcaaaaaaa aaaaaaaaaa aaaaaaaaaa aattctagta gccacattaa aaaagtataa 5640
 aagaaaaggt gaaattaatg taataataga ttttactgaa gccagcatg tccacacctc 5700
 atcattttag ggtgttatct gtgggagcat cactcacagg acatttgaca ttttttgagc 5760
 tttgtctgct ggtatccctg tgtaggctcc gtgcgtggcc atctcggcct ggacctgtct 5820
 ggtctcccat ggcctggctt gttgtaccag atggtgcagg tccgggatga ggtcggcagg 5880
 40 cctcagtgta gctggatgtg cagtgtccgg atggtgcacg tctgggatga ggtcggcagg 5940
 cctcgtctgt agctggatgt gtggtgtctg gatggtgcag gtccgggggt aggtctccag 6000
 gccctcgggt agctggaggt atggagtccg gatgatgca gtccgggggt aggtcggcag 6060
 gccctgctgt gactgtgagt tgtggtgtct ggtggtgca ggtcaggggt gaggtctcca 6120
 45 ggcctcgggt aagctggagg tatggagtcc ggtgatgca ggtccgggggt gaggtcgcca 6180
 ggcctcgtgt tgagctggat gtgtggtgtc tggatggtgc aggtctgggg tgaggtcacc 6240
 aggcctcgtg gtgagctggg tgtgcggtgt ctggatgggt caggtctgga gtgaggtcgc 6300
 cagacggtgc cagacctgc ggtgagctgg atatgcggtg tccggatggt gcaggtctgg 6360
 ggtgaggttg ccaggccctg ctgtgagttg gatgtgggtg gtccggatgc tgcaggctcc 6420
 50 gtgtgaggtc accaggccct gctgtgagct ggtatgtgtg tgtctggatg gtgaggtctc 6480
 ggggtgaaag tcgccaggcc cctgcttctg agctggatgt gtggtgtctg gatggtgag 6540
 gtctggagtg aggtcggcag gccctcgggt agctggatgt gcaggttcca gatggtgcag 6600
 gtccgggggt aggtcggcag accctcgggt gactgtgatg tgcgtgtctt ggtggtgca 6660
 ggtctggagt gaggtcgcca ggcctcgggt gactgtgatg tatggagtcc ggtggtgcca 6720
 55 ggtccgggggt gaggtcgcca gacctcgtg tgagctggat gtccgggtgtc tggatggtac 6780
 aggtctggag tgaggtcgcc agacctcgtg gtgagctgga tatgctgggt ccggatggtg 6840
 caggtcaggg gtgaggtctc caggccctcg gtgagctgga ggtatggagt ccggatggtg 6900
 caggtccggg gtgaggtctc caggccctcg tgtgaactgg atgtcggggt cctggatggt 6960
 gcaggtctcg ggtgtggtcg ccaggccctc ggtgagctgg aggtatggag tccggatgat 7020
 gcaggtccgg ggtgaggtcg ccaggccctc ctgtgagctg gatgtcgggc gtctggatgg 7080
 60 tgcaggtctg ggggtgtggtc gccaggccct cggtgagctg gaggatgga gtccggatga 7140
 tgcaggtccg ggggtgaggtc gccaggccct gctgtgagct ggtatgtctg tatccggatg 7200
 gtgcaggtccg ggggtgaggtc gccaggccct gctgtgagct ggtatgtctg tatccggatg 7260
 gtgcaggtct cgggtgaggt caccaggccc tgcgggtgag tgggtgtgct gtgtccgggt 7320
 65 gctgcaggtc cgggtgaggt tgcaggccct ctcgggtgag tggatgtgct gtgtcccggt 7380
 gctccggatg tgcaggtcca ggggtgaggt gctaggccct tgggtgggtg gatgtgccgt 7440

5 gtcaggatgg tgcaggctctg gggtagaggtc gccaggccctt tggtagagctg gatgtgcggc 7500
 gctctcatgg tgcaggctctg gggtagaggtc gccaggccctt tggtagagctg gatgtgcggc 7560
 gtcaggatgg tgcaggctccg gcttagaggtc gccaggccctt gcttagagctg gatgtgcggc 7620
 tctctggatg gtgcaggctcc gggtagaggtc agccaaggcc ttcggtagagc tggatgtggc 7680
 gtgtccggat ggtgcaggctc cggggtagagc tccaggccctt ctgaggcttag ctggatatgc 7740
 ggtgtccgga tggtagaggtc cggggtagagc gtcaccaggc cctgaggctta gctggatgtg 7800
 cgggtgtctg atggtgcagg tccggggtaga ggtcggccagg cctgctgtg agctggatgtg 7860
 gctgtatccg gatggtgcagg gtccggggtag aggtcggccag gccctgcagt gagctggatg 7920
 10 tgtgtatcc ggaagggtgca ggtctggcgt gaggtcgcca ggccctgagg ttagctggat 7980
 atgagggtg ggaagggtgca ggtcgggggt gaggtcacca ggccctgagg ttagctggat 8040
 gtgagggtg cggatgggtg aggtctgggg taggtcgcc agggccctgct gtgagctgga 8100
 tgtgctgtat ccgagtggtg cagggtccggg gtgaggtcgc caggccctgc ggtgagctgg 8160
 atgtgctgta tccggatggt gtaggtctgg cgtgaggtcg ccaggccctg cggtagagctg 8220
 gatgtgcagt gtacggatgg tgcaggctccg ggttagaggtc gccaggccctt cgggtgggct 8280
 15 gtaggtgtgt tgtctggatg gtgcaggctc gggtagagtt cggcaggccc tgcggtagagc 8340
 tggatgtgtg gtgtctggat gtcgagggtc cggggtagag tccgaggccc ctggtgagc 8400
 tggatatcg gtgtcccggt gtcgaatgg tgcaggctca gggtagagtc gccaggccctt 8460
 tggtagggctg gatgtgcccgt gtcaggatgg tgcaggctctg gggtagagtc gccaggccctt 8520
 20 tggtagagctg gatgtgcgggt gtcaggatgg tgcaggctccg gggtagagtc accaggccctt 8580
 cggtagatctg gatgtggcat gtccttctcg ttaag 8616

<210> 6
 <211> 2089
 <212> DNA
 25 <213> Homo sapiens

<400> 6
 gtactgtatc cccacgcccag gcctctgctt ctggaagtc tggaaacca gcccgccctc 60
 agcatgcccg tgtctccact tgcctgtgct tccctggctg tgcagctctg ggctgggagc 120
 30 cagggggccc gtcacaggcc tgggtccaagt ggattctgtg caaggctctg actgcttggg 180
 gctcacgttc tcttacttgt aaaatcagga gtttgtgcca agtggtctct aggggtttgt 240
 aagcagaagg gatttaaat agatggaac actaccacta gcctccttgc ctttccctgg 300
 gatgtgggtc tgattctctc tctctttttt tttcttttt tgagatggag tctcactctg 360
 35 ttgcccaggc tggagtgagc tggcataatc ttgggtcact gcaacctcca cctcctgggt 420
 ttaagcgat caccagccct agcctcctaa gtgagctggg ttacaggcac ctgcccaccac 480
 gcctgggctaa tttttgtact ttagggagag acgggggttc accatgttgg ccaggctggt 540
 ctggaactca tgacctcagg tgatccacc accctggcct cccaaagtgc tgggtttaca 600
 ggctaagcca ccgtgcccag ccccgcatct tcttttaatt catgctgttc tgaatgaac 660
 40 ttcaacttat tggatttagg tcatgagagg ataaaatccc acccacttgg cgactcactg 720
 cagggagcac ctgtgcagg agcacctggg gataggagag tccaccatg agctaacttc 780
 taggtggctg catctgaatg gctgtgagat ttgtctgca atgttcggct gatgagagt 840
 tgagattgtg acagattcaa gctggatttg catcagtgag ggacgggagc gctggtctg 900
 gagatgccag cctggctgag cccaggccat ggtattagct tctcgtgtc ccgcccaggc 960
 tgactgtgga gggctttagt cagaagatca gggcttcccc agctccccg cacactcgag 1020
 45 tccctggggg gccttgtgac acccctatgccc ccaaatcagg atgtctgagc agggagctgg 1080
 cagcagacct cgtcagaggt aacacagcct ctgggctggg gaccccgagc tgggtgctgg 1140
 gccatttccct tgcattctgg ggagggtcag ggctttccct gtgggaacaa gttaatcac 1200
 aatgcacctt acttagactt tacacgtatt taatgggtg cgacccaaca tggatcattt 1260
 accagtatct tggaaagaat ttaattgggg tgaccggaag gagcagacag acgtggtggt 1320
 50 ccccaagatg ctcttgtc caactgggac tgtgttctg cctggggggc cttggaggcc 1380
 cctcctccct ggacagggtt ccgtgccttt tctactctgc tgggctctgc gcctgagggt 1440
 agggcaccag ctccggagca ccccgggccc cagtgtccac ggagtgccag gctgtcagcc 1500
 acagatgccc aggtccagggt gtggccgctc cagccccctg gccccatgg gtggttttgg 1560
 55 gggaaaaggc caagggcaga ggtgtcagg gactgggtgg ctcatgagag ctgattctgc 1620
 tcttggctg agctgccctg agcagcctct cccgccctct ccatctgaag ggtgtggct 1680
 ctttctacct ggggttctg cctggggcca gccttgggct acccagtggt ctgtaccaga 1740
 gggacaggca tccgtgtgtg aggggcatgg gttcacgtg cccagatgc agcctgggac 1800
 caggctccct ggtgtgtgag gtgggacagt caccctgggg gttgaccgac ggactgggag 1860
 tccccagggt tgactatagg accaggtgtc cagggtgccc gcaagtagag gggctctcag 1920
 60 aggcgtctg ctggcatggg tggacgtggc cccgggcatg gccttcagcg tgtgctgccc 1980
 tgggtgccc gagccctcac tgagtcgtg ggggctgtg gcttcccggt agcttcccc 2040
 tagtctgtt tctggctgag caagcctcct gaggggctct ctattgcag 2089

10 / 18

<210> 7
<211> 687
<212> DNA
<213> Homo sapiens

5

<400> 7
gtggctgtgc tttgggttaa cttccttttt aaacagaagt gcgtttgagc cccacatttg 60
gtatcagctt agatgaaggg cccggaggag gggccacggg acacagccag ggccatggca 120
cggcgccaac ccatttgtgc gcacagttag gtggccgagg tgccggtgcc tccagaaaag 180
cagcgtgggg gtgtaggggg agctcctggg gcagggacag gctctgagga ccacaagaag 240
cagccggggc agggcctgga tgcagcacgg cccgaggtcc tggatccgtg tcctgctgtg 300
gtgcgcagcc tccgtgcgct tccgcttacg gggcccgggg accagggccac gactgccagg 360
agcccaccgg gctctgagga tccctggacct tgccccacgg ctccctgcacc ccaccctgt 420
ggctgcgggtg gctgcgggtga ccccgctcatc tgaggagagt gtggggtgag gtggacagag 480
15 gtgtggcatg aggatcccggt gtgcaacaca catgcggcca ggaaccggtt tcaaacaggg 540
tctgaggaag ctgggagggg ttctagggtcc cgggtctggg tggctgggga cactggggag 600
gggctgtctc tcccctgggt ccctatgggt ggggtgggcac ttggccggat ccactttcct 660
gactgtctcc catgctgtcc ccgccag 687

20

<210> 8
<211> 494
<212> DNA
<213> Homo sapiens

25

<400> 8
gtgggtgccg gggacccccg tgagcagccc tgctggacct tgggagtggc tgctgattg 60
gcacctcatg ttgggtggag gaggtactcc tgggtggggc gcagggagtg cagggtgacct 120
tgtcactgtt gaggacacac ctggcaccta ggggtggagg cttcagcctt tcctgcagca 180
catggggccg actgtgcacc ctgactgccc gggctcctat tcccaaggag ggtcccactg 240
30 gattccagtt tccgtcagag aaggaaccgc aacggctcag ccaccaggcc ccggtgcctt 300
gcaccccaagt cctgagccag gggctctctg tcttgaggct cagagagggg acacagcccc 360
ccctgccctt ggggtcttga gtggtggggg tcagagagag agtgggggac accgccaggc 420
caggccctga gggcagaggt gatgtctgag ttcttcgctg gccactgtca gtctctctgc 480
ctccactcac acag 494

35

<210> 9
<211> 865
<212> DNA
<213> Homo sapiens

40

<400> 9
gtaagggtta cgtgtgatag tctgtccag gatgtgtgtc tctgggatat gaatgtgtct 60
agaatgcagt cgtgtctgtg atgcgtttct gtggtggagg tacttccatg atttacacat 120
ctgtgatatg cgtgtgtggc acgtgtgtgt cgtggtgcat gtatctgtgg cgtgcataat 180
45 tgtgtgtgtg gtgtgtgtgg cacgtgtgtg tccatggtgt gtgtgcctgt ggtgtgcatg 240
tgtgtgtgtc tgtgacacgt gcatgttcat gctgtgtgtc gcatgtctgt gatgtgccta 300
tttgtgggtg gtgtgtgcat gtgtccgtga catatgcgtg tctatggcat ggggtgtgtg 360
ggcccccttg ctttactcct tcctcctcca ggcattgtcc gcaccattgt cctcacgctc 420
tcgggtgctg gtttggggag ctccacattc agggctctca cttctagcat ggggtgcccct 480
50 gtccctgtcac agggctgggc cttggagact gtaagccagg tttagagga gagtagggat 540
gctgggtgta ctttctctga cccctggcac cccagggacc ccagtctggc ctatgccggc 600
tccatgagat ataggaaggc tgattcaggc ctgcgtcccc gggacacact cctcccagag 660
cggccggggg ccttggggct cggcaggggt gaaaaggggc ctgggcttgg gttcccaccc 720
agtggtcatg agcacgctgg aggggtaagc cctcaaagtc gtgccaggcc ggggtgcaga 780
55 ggtgaagaag tatccctgga gcttcgggtc ggggagaggc acatgtggaa acccacaagg 840
acctctttct ctgactttct gagct 865

60

<210> 10
<211> 3782
<212> DNA
<213> Homo sapiens

65

<400> 10
tgtgggattg gttttcatgt gtgggatagg tgggatctg tgggattggg ttttatgagt 60
ggggtaacac agagtccaag gcgagcttcc ttctgtagt gggctctcag gtgctccaac 120

agctttattg agggagaccat atcttccctt gaactatggt cgggtttata gtaagtccagg 180
 ggtgtggagg cctcccctgg gctcccctgt ctgtttcttc cactctgggg tctgtgtggg 240
 cctgtctgtg tgtgtggcgg gtgggcaggg ctccaggccc tctctgtgtt cattggcctg 300
 gatgtggccc tggctacgct ccgtccctgg aattcccctg cgagttggag gctttctttc 360
 5 tttctttttt tctttctttt tttttttttt tgataacaga gtctcgctct tttttgcca 420
 ggctggagtg gtttggcgtg atcttggctc actgcaacct gtgcttccct agttcaagca 480
 attctcttgc ctccagcctcc caagtagctg gaattatagg cgcccaccac catgctgact 540
 aatttttgta attttagtag agacgaggtt tctccatgtt ggccaggctg gtctcgaact 600
 cctgacctca ggtgatcttc ccacctcggc ctcccaaagt gctgggagta cagggtgtga 660
 10 ccgcccggcc cgcccgagac tgccttccct cagcttccgt gagatctgca gcgatactg 720
 cctgacgctt tgggtgctgac aaacctcgtt tctcttctcc aggtctcgtt aggggtcttt 780
 ccatttcatg actctcttca cagaagagtt tcacgtgtgc tgatttccc gctgtttcct 840
 gcgtaattgg tgtctgctgt ttatcgatgg cctccctcca tttcctttag gctttgttta 900
 ttgtgtttt tccggctcct tgaaggaaaa gtctcgatta tggatgtttg aactttcttt 960
 15 tctaaacaag catctgaagt tgccgttttc cctctaaagc agggatcccc aggccctgg 1020
 ctgtggagtg gcaccggtct ggggctgtt aggaaccggc cgcacagcgg gaggctagg 1080
 ggggtgtggg gagccagcgt tcccgctga gccccgcccc tctcagatca gcagtggcat 1140
 gcgggtgctca gaggcgaca caccctactg agaactgtgc gtgagagggg tctagattct 1200
 20 tgcctcctta tgggaatcta atcgctgatg atctgaggtg gaaccggttg ctcccaaac 1260
 catccccctt cccactgctg tctgtggaa aaatcgtctt ccacgaaacc agtccctgg 1320
 accacaatgg ttggggaccc tgtgtctaaag acctgcttca gcagcctctc gtcagtgtt 1380
 atatatggc tttctgtgtt tgagctccaga ataattacgg atttctgtga tgccttccg 1440
 cgacctcaga cccatgggct atttgtgggc gtgttgcctg ctctgggtt gggaaaggg 1500
 25 caggccccc atgaccttct gtactgctt tccagggttg tctcagggt tgaatcgtac 1560
 tctgatgtgt ttagccac ggccttgcg ccagctcctg ggggctggg aacatgctga 1620
 agcacagagt caccgtgcgc gtcttctgat gctcacaag ctcgaggcct cctgtgtccg 1680
 tgttagtgtg tgtcacgtgc ctgctcacat cctgtcttgg ggacgcaggg gcttagcagg 1740
 tcccgtagta aatgacaagc gtccctgggg agtctgcaga ataggaggtg ggggtgccc 1800
 30 tctctctccc gcgtcttcag actcttctcc tgcctgtgct gtggctgcac ctgcatcct 1860
 gcaatccctc cagcactggg ctggagaggg ccgggagctc gaggccact tgtgccactg 1920
 gactgtggat ggcagtcggc caggggggtc tgatgtgtgg tgactgtgga tggcgggtg 1980
 tcacaggggt ctgatgtgtg gtgactgtgg atggcggtcg tggggctcta tgtgggtact 2040
 35 tgggatggcg gtctgtgggt ctgatgtgtg gtgactgtgg atggcggtcg tggggctcta 2100
 tgtggtgact gtggatggcg gtctgtgggt ctgatgtgtg gactgtggat ggcggctcgt 2160
 gggctctgat tgggtgactgt ggtggcagc cgtgggggtc gatgtgtgtg gactgtggat 2220
 ggcggctcgt gggctctgat tgggtgactgt ggtggcagc cgtgggggtc gatgtgtgtg gactgtggat 2280
 gactgtggat ggcggctcgt gggctctgat tgtgtgact gtggatggcg gctgtgggt 2340
 ctgatgtgtg gtgactgtgg atggcggtcg tggggctcta tgtgtgtga ctgtggatg 2400
 40 cggctcgtgg gtctgatgtg gtgactgtgg atggcggtcg tggggctcta tgtgtgtga 2460
 ctgtggatg tgatcggtca caggggtctg atgtgtgtg actgtggatg gcggtcgtg 2520
 ggtctgatgt gtggtgactg tggatgggtga tgggtcacag gggctctgat tgtgtgact 2580
 gtggatggcg gtctgtgggt ctgatgtgtg gtgactgtgg atggcggtg gtcccgggg 2640
 tctgatgtgt ggtgactgtg gatggcgatc ggtcacaggg gtctgatgtg tgggtgactg 2700
 45 ggtggcggt cgtgggggtc gatgtgtgtg gactgtggat ggcggctcgt gggctctgat 2760
 tgtggtgact gtggatggcg gtctgtgggt ctgatgtgtg gactgtggat ggcggctcgt 2820
 gggctctgat tgggtgactgt ggtggcggt cgtgggggtc gatgtgtgtg gactgtggat 2880
 ggcgggtgtt cccgggggtc tgatgtgtg tgactgtgga tggcggtcgt ggggtctgat 2940
 gtggtgactg tggatggcag tctgggggtc tgatgtgtg tgactgtgga tggcggtcgt 3000
 50 ggggtctgat gtgtgggtgac tgggtatggc ggtcgtgggg tctgatgtgt ggtgactgtg 3060
 gatggcggtc gtggggctc atgtgtgtg actgtggatg gcggtcgtg ggtctgatg 3120
 ggtgactgtg gatggcggtc gtggggctc atgtgtgtg actgtggatg gtgatcggtc 3180
 acaggggtct gatgtgtgtg gactgtggat ggcggctcgt gggctctgat tgtgtgact 3240
 gtggatggcg gtctgtgggt ctgatgtgtg gactgtggat ggcggctcgt gggctctgat 3300
 55 tgtgtgact gtggatggcg gtctgtgggt ctgatgtgtg gtgactgtgg atggcggtc 3360
 gtcacagggg tctgatgtgt ggtgactgtg gatggcggtc gtggggctc atgtgtgtg 3420
 actgtggatg ggcgtcgtg ggtctgatgt gtggtgactg tggatggcg tctgtgggtc 3480
 tgatgtgtg tgactgtgga tggcggtcgt ggggtctgat tgggtgactg tggatgggtg 3540
 tgggtcacag ggtctctgat tgtgtgtg gtaggtggag tccaggtgt gctctgact 3600
 60 acttctgctc ctgcggcccc cgccccctt tcccaaca gaagcttccc aggcgtctc 3660
 tgggtctcat cccgccatcg ggtctggcgg caggtccaca cgtcctgat ggaagaaaca 3720
 agtggccagc tctggccggg gcaggccaca tctgtggctc atgcctctc ctctgcggcg 3780
 ag 3782

12 / 18

<210> 11
<211> 980
<212> DNA
<213> Homo sapiens

5

<400> 11
gtctgggcac tgccttgacg gggtgggcac ggactcccag cagtgggtcc tcccctgggc 60
aatcactggg ctcatgaccg gacagactgt tggccctggg gggcagtggg gggaaatgagc 120
10 tgtgatgggg gcatgatgag ctgtgtgcct tggcgaaatc tgagctgggc catgccaggc 180
tgccagacgt gctgcattca ggcacctgct cacttttgac tgcgcggcct ctctccagtt 240
ccgcagtggc tttgttcatg atttgcataa tgtcttctct gccagttttg atcttgaggc 300
caaaggaaaag gtgtcccccct ccttttaggag ggcaggccat gtttgagccg tgtcctgccc 360
agctggcccc tcagtgtctg gtctgaggcc aaaggaaacg tgtcccccct cttaggaggga 420
cggggcgtgt ttgagccacg ccccgtgag cgggcctctc agtgcctggg ctgtccacgt 480
15 ggcctgttg ccctttgcag atgtggtctg tccacgtggc cctgtggctc tttgcagatg 540
ctgttttagc cttgtctggc tctaggggac agtgcgtgtcc accgcattgag gctcagagac 600
ctctggggca atttcccttg ctcccagggt gggggtggag gtggcctggg ctgctgggac 660
ccagaccctg tgcctggcag ctgggcagca actcctggat cacatatgcc atccggggcca 720
gggtgggtgt tgtgggtgtg agcccagctg gacccacagg tggcccagag gagacgttct 780
20 gtgtcacaca ctctgcctaa gcccatgtgt gtctgcagag actcggcccc gccagccccc 840
gatggccctg cattccagcc cagccccgca ctctatcaca aacactgacc ccaaaaggga 900
cggagggtct tggccacgtg gtccctgctg tctcagcacc caccggctca ctcccatgtg 960
tctccgtctc gcttttcgag 980

25

<210> 12
<211> 2485
<212> DNA
<213> Homo sapiens

30

<400> 12
gtgagtcagg tggccagggt ccattgccct gcgggtggct gggcgggctg gcaggggcttc 60
tgctcacctc tctcctgccc ctccccact gncctttctg ccggggggcc cagagctctcc 120
ttttctggcc ccgcgccccc cgggctcctg ggctgcaggc tcccaggagg ccggaaacat 180
ggctcggctt gcggcagccg gagcggagca ggtgccacac gaggcctgga aatggcaagc 240
35 ggggtgtgga gttgtcctct cgtggaggac gaggggcggg ggggtgtgtc gggtcagggt 300
tgcgcggagc gtttgagcct gcagcttgtc agctccaagt tactactgac gctggacacc 360
cggctctcac acgcttgtat ctctctctcc cgatacaaaa ggattttatc cgatttctat 420
tctgtccct gtctgtgtac ccccgcgagg gcgcgggctc tctctctgt gactagattt 480
cccatctgga aagtgcgggg ttgaccgtgt agtttctctc tctcgggggg cctgtgggtg 540
40 ccatggggca ggcggccttg gagagctgcc gtcacacagc cactgggtga gccacactca 600
cgggtggtaga gccacagtgc ctggtgccac atcacgtcct ctggatttta agtaaaacca 660
cacacctccc ggcaggcatc tgccctgcgac cctgtgtgtg cctggggaga gtggtagcac 720
ggaggaaatt cgtgcacatc caaggtcatc agcaaggta tccgcagtca ggtggaacct 780
ggaggcctct ctctgggtgc gtctccagcg gataaaggac tgtgcacagc ttcggaaagt 840
45 tttatttaaa aatataacta ttaattattg cattataagt aatcactaat ggtatcagca 900
attataatat ttattaaagt ataattagaa atattaaagta gtacacacgt tctggaaaaa 960
cacaaattgc acatggcagc agagtgaatt ttggccgagg gacacgtgtg cacatgtgtg 1020
taagcggccc ccaggccccc agaattcgct gacaaagtca cctccccaga gaagccacca 1080
cgggcctcct tctgtgtcgt gaattttatt aagatggatc aagtcacgta ccgtccacgt 1140
50 gtggcaggggc tttggggaat gtgaggtgat gactgcgtcc tcatgccctg acagacagga 1200
ggtgactgtg tctgtcctgt ccctaggaca cggacaggcc cgaagctcta gtccccatcg 1260
tggctccagtt tggcctctga ataaaaacgt cttaaaaaac tgttgcccca aaaactaaga 1320
acagagagag tttcccatcc catgtgtctca caggggcgta tctgcttgcg ttgactcgct 1380
gggctggccc gactcctaga gttggtgct gtgtctctgt gcaaaaagtg cagtcctctt 1440
55 gcccatcact gtgatatctg caccagcaag gaaagcctct tttcttttct tctttttttt 1500
ttttttgaga cggaaacgtc ctgtgtctct cctgggcttg agtgcagtgg cgcgatctca 1560
actcactgca acctccgcct cccgggttcc agcattttct ctgcctcagc ctcccagca 1620
gctgagatta caggcaccca ccccctgcgc ctggctaatt tttgtatttt tagtagagag 1680
gggtttttgc catgtttggc aggctggtct cgaactcctg acctcagggt atccaccac 1740
60 ctccgctctc caaagtgtg ggattacagg tgtgagccat cagccccagc cggaaagcct 1800
ctttttaagg tgaccaccta tagcgcttcc cgaaaataac aggtcttctt tttgcagtga 1860
gctgcaagcg tctcttagca acaggagtgg cgtcctgtgg gctctgggga tggctgaggg 1920
tcgcgtggca gccatgcctt ctgtgtgcac cttaggttcc cagggggcta tctgtctctc 1980
actgtttgtc tgaaaacgca ccttggcat ccttgtttgg agagtctctg ctctcgtctg 2040
65 gtcattgtga aactaggggc aagggtgcat ccgttggtgc gcagcggcta catgtagggt 2100

13 / 18

catgagtcctt tcaccgtgga caaattcctt gaaaaaaaaa aaaggagtcg gggttaagcat 2160
 tcattccggg tcaagtgtct ggttctgtga ataaactcta agatttaaga aaccttaatg 2220
 aaagaaaacc ttgatgattc agagcaaggga tgtggtcaca cctgtggctg gatctgtttc 2280
 agccgccccca gtgcatggtg agagtgggga gcagggttg tttgttcaga ggtctcatct 2340
 5 ggtatgtttc tgagggtgtt gccggctgaa tggtagacgt gtcgtttgtg tgtatgaggt 2400
 tctgtgtctg tgtgtggctc ggtttgagtg tacgcatgtc cagcacatgc cctgcccgtc 2460
 tccacactgt gtcttcccgc cccag 2485

<210> 13
 <211> 1984
 <212> DNA
 <213> Homo sapiens

<400> 13
 15 gtgaggcctc ctcttcccca ggggggcttg ggtgggggtt gatttgcttt tgatgcattc 60
 agtggttaata ttctctgtgc tctggagacc atgactgttc tgtcttgagg aaccagacaa 120
 ggttgacagc ccttcttggg atgaagccgc acgggagggg ttgcacagcc tgaggactgc 180
 gggctccacg caggctctgt ccagcggcca tgtccagagg cctcagggtc cagcaggcgg 240
 gagggccgct gccctgcacg atgagcatgt gaattcaaca ccgaggaagc acaccagctt 300
 20 ctgtcacgtc acccaggttc cgttaggggc cttggggaga tggggctggg gcagcctgag 360
 gccccacatc tcccagcagg ccctcgacag gtggcctgga ctgggcgcct cttcagccca 420
 ttgcccaccc cacttgcacg ggtctacac ccaaggacgc acacacctaa atatcgtgcc 480
 aacctaatgt ggttcaactc agctggcttt tattgacagc agttactttt ttttttttaa 540
 tactttaagt tctagggtag atgtgcacga cgtgcagggt agttacatat gtatacatgt 600
 gccatgttgg tgtgctgcac ccattaactc atcatttaca ttaggtatat ctcttaatgc 660
 25 tatccctccc cactccccc atcccatgac aggcctctgt gtgtgatgtt cccaccctg 720
 tgtccaagtg ttctcattgt tcagttccca cctgtgagtg agaacatgtg gtgtttgggt 780
 ttctttcctt gcaatagttt gctcagagtg atggtttcca gcttcgtcca tgtccctaca 840
 aaggacatga actcatcctt ttttatgact gcatagtatt ccgtgggtga tatgtgccac 900
 30 attttcttaa tccagctctat catcgatgga catttgggtt ggttgcaagt ctttctact 960
 gtgaatagtg ccgcaataaa catacgtgtg catgtgtctt tatagcagca tgatttataa 1020
 tcctttgggt atataccacg taatgggatg gctgggtcaa atggtatttc tagttctaga 1080
 tccttgagga atccacacac tgtcttccac aatggttgaa ctagtttaca ctcccaccaa 1140
 cagtgtaaaa gtgttctggt gctggagagg atgtggacag cagttatttt tttatgaaaa 1200
 35 tagtatcact gaacaagcag acagttagtg aaggatgctt cagggaagcct gcaggccaca 1260
 cagccatttc tctcgaagac tccgggtttt tctgtgcat cttttgaaac tctagctcca 1320
 attatagcat gtacagtgga tcaaggttct tcttcattaa ggttcaagtt ctagattgaa 1380
 ataagtttat gtaacagaaa caaaaatttc ttgtacacac aacttgctct gggatttgga 1440
 ggaaagtgtc ctgcagctgg cggcacactg gtccagccctc tgggacagga tacctctggc 1500
 40 ccattggtcat ggggcgctgg gcttggccct gaggggtcaca cagtgcacca tgcagcgtt 1560
 cctgtggata ggatctgggt ctcgatcat gctgaggacc acagctgcca tgcgtgtaaa 1620
 gggcaccacg tggctcagag ggggcgaggt tcccagcccc agctttctta ccgtcttcag 1680
 ttatttttcc ctaagagtct gagaagtggg gccgcgctg atggccttcg ttcgtcttca 1740
 45 gctggcacag aattgcacaa gctgatggta aacactgagt acttataatg aatgaggaat 1800
 tgctgtagca gtttaactgt gagagctcgt ctgttggaaa gaaatttaag tttttcattt 1860
 aaccgctttg gagaatgtta ctttatttat ggtgtgtgaa attgtttgac attcagctccc 1920
 tcgtagacag atactacgta aaaagtgtaa agttaacctt gctgtgtatt ttcccttatt 1980
 ttag 1984

<210> 14
 <211> 1871
 <212> DNA
 <213> Homo sapiens

<400> 14
 55 gtgaggcccg tgccgtgtgt ctgtggggac ctccacagcc tgtgggcttt gcagttgagc 60
 ccccggtgtc ctgccccctg caccgcagcg ttgtctctgc caagtcctct ctctctgccg 120
 gtgctggatc cgcaagagca gaggcgcttg gccgtgcacc caggcctggg ggcgcagggg 180
 60 caccctcggg agggagtggg tacctgtcag gccctgttcc tgcagagacg caccaggtt 240
 acacacgtgg tgagtgcagg cgtgtacctg gctcctgtct ctctttggaa agtcaagagt 300
 ggcggtcctt ggggccccag tgagacccc agagctgtg cacagggcct gcagggccga 360
 ggccggcagcc tcctccccag ggtgcacctg agcctgcgga gagcaggagc tgcgtagtga 420
 gctggcccac agcgttctgt gcggtcacgt tcctgcgttg ggttgttttg gatcggtggg 480
 agaatttggg tttgtgagt gctgctgtct tgaaccacg agatggctag gagtgggttt 540
 65 cagagttgat ttttgtgaat caaaactaaa tcaggcacag gggacctggc ctccagcacg 600

14 / 18

5 gggattgtcc aatgtggtcc ccctcaaggg cgccccacag agccgggtggg cttgttttaa 660
 agtgcgattt gacgaggggac gagaacaccc gaaagctgta aaggggaaccc tcagaaaatg 720
 tggccgcagc ggggtggttc aggtgctttg ctgggctgtg tttgtgaaa cccatttgga 780
 cccgccctcc aagtcacccc tccaggtcca cctccaggg ccgcctggg ctgggggtat 840
 10 gcctggcggt ccttgtgccc cagcccgag cacagcaggc tgtgcacatt taaatccact 900
 aagattcact cggggggagc ccaggtccca agcaactgag ggctcaggag tccagaggct 960
 gctgagggga cagagcagac ggggaacgct gcttctgtgt ggcaagtcc tgaggggtgt 1020
 ggccaggagg gtggctcaga gtgtatgttg gggccccacc gggggcagaa ctctgtctct 1080
 gatgagtcgg cagccatgta acaggaaggg gtggccacag ggagctggga atgcaccagg 1140
 ggagctgcgc agctggccga ggtcccagg ccaggccaca ggaaggcagc ggggacgccc 1200
 gggggccacag cagagggcgc aggaaggga ggggatgccc aggccagagc agaggctacc 1260
 gggcacaggg gggctccctg agctgggtga gcgaggctca tgactcggcg agggaaacct 1320
 cttgacgtga agctgacgac tgggtgttgc cagctcacag cccagccagg tcccgcgcct 1380
 15 gagcaggaac tcagaaacct ccccttctgc taaagcacag cagatgcctt cagggcatct 1440
 aggaagaaac agggaaagtc gttgagaaac gtcttaaaag aaggtgggat ggtggcaatt 1500
 tcttctccag atttttagtct gccccggacc acagatgagt ctataacggg attgtggtgt 1560
 tgccatgggg acacatgaga tggaccatca cagaggccac tggggctgca cctcccatct 1620
 gagtccctgc tgtcccgggt ccaggccagg ttcttgcatg ctacactacc tgtccctgcc 1680
 20 gggagacagg gaaagcacc cgaagtctgg agcagggtg ggtccaggct cctcagagct 1740
 cctgccaggc ccagcacctt gctccaaatc accactctc tgggggtttc caaagcattt 1800
 aacaagggtg tcaggttacc tcttgggtga cggccccgca tcttggggct gacattgccc 1860
 ctctgcctta g 1871

25 <210> 15
 <211> 3801
 <212> DNA
 <213> Homo sapiens

30 <400> 15
 gtgagcgac ctggccggaa gtggagcctg tgcccggtg gggcagggtg tgctgcaggg 60
 cegtgtgcgc caccctctgt tccgtgtggg gcaggcgact gccaatccca aagggtcaga 120
 ggccacaggg tgccctctgt cccatctggg gctgagcaga aatgcattct tctgtgggag 180
 tgaggggtgt cacaacggga gcagttttct gtgtattttt ggtaaaagga aatggtgcac 240
 35 cagacctggg tgacttgagg tgtcttcaga aagcagctg gatccgaacc caagacgccc 300
 gggccctgct gggcgtgagt ctctcaaac cgaacacagg gggcctgctg ggcagagtc 360
 cctctgaacc cgagaccctg gggccctgct gggcgtgagt ctctccgaac ccagagactt 420
 cagggccctt ttgggcgtga gtctctccgc tgtgagcccc acactccaag gctcatccac 480
 agtctacagg atgccaatgag ttcatgatca cgtgtgaccc atcaggggac agggccatgg 540
 40 tgttgggggg gtctctacaa aattctgggg tcttgtttcc ccagagcccc agagctcaag 600
 gccccgtctc aggtcagac acaaatgaat tgaagatgga cacagatgca gaaatctgtg 660
 ctgtttcttt tatgaataaa aagtatcaac attccaggca gggcaagggt gctcacacct 720
 ataatcccag cactttggga ggccgaggtg ggtggatcac ttgagggcag gagtttgagg 780
 ccaacctaac caacatagtg aaattccatt tctacttaaa aaatacaaaa attagcctgg 840
 45 cctggtggca cagcctgta gtccccgcta tgccggaggc tgaggcagga gaattattt 900
 aacccaggag gcagaggttg cagttagccg agatcacacc actgcactcc agcctgggca 960
 acagagttag acttcatctt aaaaaaaaaa aaaaaagtat cagcattcca aaaccatagt 1020
 ggacagggtt ttttttattc tgtccttcca taatatctac tgggtgctgtg ctgagggccc 1080
 gaactggggg tgccctctct tgaaggacac accttcatgg gaagagaaat aagtgggtga 1140
 50 tgggtgttaa accagaggtt taaactgggg tctgtctgt ctgagttaac agtccagatc 1200
 tggactttgc ctctttccag aatgctccct ggggtttgct tcatggggga gcagcaggtg 1260
 tggacacccc cgtgatgggg gagcagcagg tgcagacgac ctcatgatgg gggagtggca 1320
 ggtgcagaca ccttctgcca tgggtgccag catgtccctg ttgcagctcc ctccccacaa 1380
 ggatgccggt ctctctgtgt cccacagtc cctgcttccc tctcacagcc ttacctgggc 1440
 55 ctggcctcca ctggctttgt ctgcatgatt tccacatttc ctgggctccc agcactctct 1500
 cgctctctcc aggcacctct gcagtgtctg ccataccagt cagctgtgaa ctgtccactg 1560
 cttattttgc tccccatgaa atgtattttt taggacaggc acccctggtt ccagcctctg 1620
 gcacagcatc agtgaatggt attgaaggac aaaggacaga caaacaatc aggaataatg 1680
 gttctctcta aacacattgc aaagccacag aggttagtgc aggatgggtg ggcacagggt 1740
 60 catcagatgt ggggtccaatg ccagaatatt ctgtgctccc aaaggccact tggtcagagt 1800
 gtgtgcttgc agaggtgggt ctaaaagctc agcagtggag gcagtgggtc gccatactca 1860
 ggggtgaactc acatctctgt tgtctgaagt atacagcaga ggcttgaagg gcatctggga 1920
 gaagaaaaca ggcaaaatga ttaagaaaag tgaaaaagga aaagtggtaa gatgggaatt 1980
 ttcttgccta gatttttagtc tcccaaacca cagctcagat ggtagaatgt ggtcagaact 2040
 65 gatggacaga acaatagaac aaaacggaag ccttatctct cagaaacgtg tgttaattgt 2100
 gtatgtggca cagctgatgg aaaagagagt gtgtgtgtaa tttttttttc tgagaaaact 2160

15 / 18

gactggaagc aaataagttg tgtctttaca gcatatacca gagcagattc taggtagaag 2220
aggagacaca tgcaaacac accagcaaca gaaataaac aaaagactca aaggggaagg 2280
agggtgaacgt tccctgggtt ggtgttgggg aaggacacac agggaggcgg atgaaaccag 2340
5 tgaggcaacg ggcattgctt tctactgcaga gaaactcagc ttgcctgagc cacagtgaag 2400
atggccattc cctggagcgt ttgtgcacgt gatttattta aggcgcctcg tgaggctctg 2460
cacattcatc ctctcacttt gtctcctaa ccacctgaga ggtagaggag gaaaggctcc 2520
aggggagcag cgcctcttgg tcacccagct ggcaaggggc atgcatgatt gcagcctggc 2580
ctcctgctcc ggggcccctg ctctgccga ggacccaca caagtcagac ccataggctc 2640
10 aggttgagcc ggagcccaag gtcgtgttgg ggatggctgt gaaagaagaa atggacgtct 2700
gatgcacact tgggaaggtc ctaccagcag cgtcaaagaa atgcatgtga aactgacagc 2760
gagaccatc cctcaaagaa acgcacgtga aactgatggc gagacctgtc cccatccctc 2820
atgctggctc cttttctggg cttgccaaaga gccagcatca ggttgaggca agctggaaag 2880
acttttctgg aaagcagctt gtttgcattg aagtcctcac aatgtcctgt gtcttccag 2940
taattccact tctgaagtga ccagacatta tcacgggtct tatttaccat ttccagtggt 3000
15 ccaggcaggg ggacttgcca cagcaagtca cgaacctgcc caaatacagg gctaaggaga 3060
tattatgcat cacaaaactt gctctgccat taaacathtt tcaaagaatt ttggaagaat 3120
gtttaatggc acaaaacgtt tatttcaatg tagcagtggt caaagctgga tgtaaaagaa 3180
cacacccag gagcctgccg tgaatgtcat gtgtgttcat ctttggacat ggacatacat 3240
gggcagtga tggtgtgag gccctggagg acatcggtgg gatgcctcca tcctgcccct 3300
20 ctggagacac catgtgtgcc acgtgcactc actggagccc tgtttagctg gtgccacctg 3360
gctcttccat ccctgagatt caaacacagt gagattcccc acgcccactc cagtgttctc 3420
ccacaaaaaa cctgagtcac acctgtgttc actcgaggga cggccgggag ccagggtctc 3480
acagtttatt atgtgttttt ggctgagtta tgtgcagatc tcatcagggc agatgatgag 3540
tgcaaaaaa cggcgtgtcg aggtttggat acactcaaca tctactagcca ggtcctggtg 3600
25 gaggtttggtc atgcagagtc tggatggcat gtagcatttg gagtccatgg agtgagcacc 3660
cagcccccctc gggctgcagc gcatgcccga ggagggacaa ggaagcggga ggaaggcagg 3720
aggctctttg gagcaagctt tgcaggaggg ggctgggtgt ggggcaggca cctgtgtctg 3780
acattccccc ctgtgtctca g 3801

30 <210> 16
<211> 880
<212> DNA
<213> Homo sapiens

35 <400> 16
gtgagcaggc tgatgggtcag cacagagttc agagtccagg aggtgtgtgc gcaagtatgt 60
gtgtgtgtgt gtgcgcgcgt gccctgcaagg ctgatggtga ctggctgcac gtaagatgtc 120
acatgtacgc atatacacgt gagcacatac atgtgtgcat gtgtgtacat gaaggcatgg 180
cagtgtgtgc acagggtgtgc aaggggacaa gtgtgtgcac atgcgaatgc acacctgaca 240
40 tgcattgtgt ttctgtgcaca gtctgtgtgg cattcacgtg aggtgcatgc gtgtgggtgt 300
gcagtgtgag tagcatgtgt gcacataaca tgtattgagg ggtcctcgtg ttcaccccgc 360
taggtcctca gcaccagtgc cactccttac aggatgagac ggggtcccag gccttggtgg 420
gctgaggctc tgaagtgcga gccctgaggg cattgtccca tctgggcac cgcgtccact 480
50 cctctcctg tgggcttctg tgtccactcc cctctcctg tgggcattta catccactcc 540
45 actccctctc tctctgtggc atccgcgtcc actcccccctc tctgtgggca tctgcgtcca 600
cctcccccctc ctgtgggcat ttgcgtccac tccctctcct gggtccttcc tgtcttggcc 660
gagcctcggg ggcaggcaga tgacacagag tcttgactcg ccaggggtgg ttcgcagctg 720
ccgggtgagg gccaggcccg atttactcgg gaagagggat agtttcttgt caaaatgttc 780
50 cttcttcttg ttccatctga atggatgata aagcaaaaag taaaaactta aaatcccaga 840
gaggtttcta ccgtttctca ctcttctctg gcgactctag 880

55 <210> 17
<211> 3186
<212> DNA
<213> Homo sapiens

60 <400> 17
gtgagccgcc accaaggggg gcaggcccag cctccaggga ccctccgcgc tctgtctacc 60
tctgaccggg ggcttcacct tggaactcct ggggttttagg ggcaagggaat gtcttacgct 120
ttcagtggtg ctgctgcctg tgcacagttc tgttcgcgtg gctctgtgca aagcacctgt 180
60 tctccatctc tgggtagtgg taggagccgg tgtggcccca ggtgtcccca ctgtcctgt 240
gcactggccg tgggacgtca tggaggccat cccagggcag caggggcatg gggtaaagag 300
atgtttatgg ggaagtcttag cagaggaggg tgggaagggt tctgaacagt agatgggaga 360
tcagatgccg ggaggatttg gggctctcagc aaagagggcc gaggtgggtg caggtgaggg 420
65 tgcgtggccc cacccccggg aagggtgcagc agagctgtgg ctccccacac agcccggcca 480

16 / 18

5 gcacctgtgc tctgggcatg gctgtgctcc tggaaacgttc cctgtcctgg ctggtcaggg 540
 ggtgccccctg ccaagaatcg acaactttat cacagaggga agggccaatc tgtggaggcc 600
 acaggggccag cttctgcctg gagttagggc aggtgggtggc acaagcctcg gggctgtacc 660
 aaagggcagt cggggcaccac agggccgggc ctccacctca acaggcctcc cgagccactg 720
 ggagctgaat gccaggaggc cgaagccctc gcccctagag ggctgagaag gagtgtgagc 780
 atttgtgtta ccaggggcgg aggtgcgcgc aattaccgtg cacacttgat gtgaaatgag 840
 gtgcgtgctc atcgtggaaa ccagcgaagg gctcacggga gagttttcca ttacaaggtc 900
 gtacccatgaa aatgggtttt aaaccgagtg cttgcgcctt catgctctgg cagggagggc 960
 10 agagccacag ctgcatgtta ccgcctttgc accagctcca gaggcttggg accaggctgt 1020
 ctcagttcca ggggtgcgtcc ggctcagacc gccctcctct ctgctctctc tctctgcctc 1080
 aaatcttccc tcgtttgcat ctccctgacg cgtgcctggg ccctcgtgca agctgcttga 1140
 ctcccttccg gaaacccttg ggggtgtgctg gatacaggtg ccactgagga ctggagggtg 1200
 ctgacactgt ggttgacccc aggggtccagc tggcgtgctt ggggcctcct tgggccatga 1260
 tgaggtcaga gtagttttcc cagggtgaaa ctccctgggaa actccaggg ccagtgtacc 1320
 15 tggccactgc tctctccata ttcagctcag tcttgcctc atttcccac cagggtctctc 1380
 agctccgagg agctcccgtg aggggccttg gctcaggga gggcggtga gtttcccac 1440
 ccatgtgggg acccttgggt agtcgcttga ttgggtagcc ctgaggaggc cgagatgcga 1500
 tggggcacgg gccgtttcca aacacagagt caggcacgtg gaaggccag gaatcccct 1560
 20 cctcgaggc aggagtgagg gaacggagag ctgggccccg atttcacggc agccaggctg 1620
 cagtgggcga ggctgtggtg gtccacgtgg cgtggggggc ggggtctgat tcaaatccgc 1680
 tggggctcgg ccttctctgg ccgtgctggc cgcgcctcca cacgggcttg ggggtggacg 1740
 ccgcacctct agcaggtggc tatttctccc ttggaagag agccctcac ccagtctag 1800
 tgttctccctc ctgggtcagg agcgtggccg tgtggcaacc ccgggacctt aggcctattt 1860
 25 atttgtttaa aaacattctg ggcctggctt ccgttgttgc taaatgggga aaagacatcc 1920
 cacctcagca gagtactga gaggctgaaa ccgggggtgt ggcttgactg gtgtgatctc 1980
 aggtcattcc agaagtggct cagggaagtca gtgagaccag gtacatgggg ggctcaggca 2040
 gtgggtgaga tgaggtacac ggggggctca ggcagtggtt gaggccaggt acatgggggg 2100
 ctcaggcact ggtgtgagat aggtacacgg ggggctcagg cagagggtca gaccagggtac 2160
 30 acgggggctc tgatcacacg cacatatgag cacatgtgca catgtgctgt ttcattgtag 2220
 ccaggctctgt gcacacctgc cccaaagtcc cagggaagctg agaggccaaa gatggaggct 2280
 gacagggctg gcgcgtggc tcacacctgt agtcccagca ctttgggagg ccgaggcgag 2340
 aggatccctt gagcccagga gtttaagacc agcctgagca acatagtaga accccatctc 2400
 35 tatgaaaaat aaaaacaaaa attagctgaa catggtggtg tgcgcctgta gttccaatac 2460
 ttgggaggct gaagtgggag gatcacttga gcccgaggag tggaaagctg agtgagctga 2520
 gattgcacca ctgtactgca gcctgggtga cagagtgaga gcccatctca acaacaacaa 2580
 agaagactga caaatgcagt ttcttggaaa gaaacattta gtaggaaactt aacctacaca 2640
 cagaagccaa gtccgtgtct cgggtgcagt gagatgagat gatgggtcct cacaccatca 2700
 ccccgagccc aggggtttatg caccacaggg gcgggtggct cagaagggat gcgcaggacg 2760
 40 ttgatatacg atgacatcaa ggttgtctga cgaaggcgag gattcatgat aagtacctgc 2820
 tgggtacaaa ggaacaatgg ataaactgga aaccttagag gccttcccgg aacaggggct 2880
 aatcagaagc cagcatgggg ggctggcatc caggatggag ctgcttcagc ctccacatgc 2940
 gtgttcatac agatgggtga cagaaacgca gtgtacctgt gcacacacag acacgcagct 3000
 actcgcacac acaagcacac acacagacat gcatgcatgc atccgtgtgt gtgcacctgt 3060
 45 gcccatgagg aaaccatgc atgtgcattc atgcacgcac acaggccacc gtggggccat 3120
 gcccacacc acgagcaccg tctgattagg aggcctttcc tctgacgtg tccgccatcc 3180
 tctcag 3186

<210> 18
 <211> 781
 50 <212> DNA
 <213> Homo sapiens

<400> 18
 55 gtatgtgcag gtgcctggcc tcagtggcag cagtgcctgc ctgctggtgt tagtgtgtca 60
 ggagactgag tgaatctggg cttagggaagt tcttaccctt tttcgcatca ggaagtgtt 120
 taaccaaac actgtcaggc tcgtctgccc gccctctcgt ggggtgagca gagcacctga 180
 tggaaaggac aggagctgtc tgggagctgc catccttccc acctgtctct gcctggggaa 240
 gcgctggggg gcctgggtctc tctgttttgc cccatgggtg gatgtggggg gcctggcctc 300
 60 tcctgtttgc cctgtggtgg gattgggctg tctcccgtcc atggcactta gggcccttgt 360
 gcaaaaccag gccaaaggct taggaggagg ccaggccacc gctacccac ccctctcagg 420
 agcagaggcc gcgtatcacc acgacagagc cccgcgcgt cctctgtctc ccagtcaccg 480
 tcctctgccc ctggacactt tgtccagcat caggagggtt tctgatccgt ctgaaattca 540
 agccatgtcg aacctgcggt cctgagctta acagcttcta ctttctgttc tttctgtgtt 600

17 / 18

gtggaaattt cacctggaga agccgaagaa aacatttctg tctgtactcc tgcggtgctt 660
gggtcgggac agccagagat ggagccaccc cgcagaccgt cgggtgtggg cagctttccg 720
gtgtctcttg ggaggggagc tgggtctggg cctgtactcc tcagcctctg ttttccccc 780
g 781

5
<210> 19
<211> 536
<212> DNA
<213> Homo sapiens

10
<400> 19
gcaagtgtgg gtggaggcca gtgcgggccc cacctgccca ggggtcatcc ttgaacgcc 60
tgtgtggggc gagcagcctc agatgtctgt gaagtgcaga cgccccggg cctgaccctg 120
ggggcctgga gccacgctgg cagccctatg tgattaaacg ctggtgtccc caggccacgg 180
15 agcctggcag ggtccccaac ttcttgaacc cctgcttccc atctcagggg cgtatggctcc 240
ccacgcttgg gagccttctg acccctgacc tgtgtctctc cacagcctct tccctggctg 300
ctgccctgag ctcttggggg cctgagcaag ttcttctccc gccccggcgc tccagcgtca 360
ctgggtctgcc tgtctgtctg ccccggtgga ggggtgtctg tcccttcaat gaggttccca 420
ccagccaggg ccacgaggtg caggccctgc ctgccgggc acccacacgt cctaggaggg 480
20 ttggaggatg ccacctctgg cctcttcttg aacggagtct gattttggcc ccgcag 536

<210> 20
<211> 3179
<212> DNA
25 <213> Homo sapiens

<400> 20
atctcatgtt tgaatcctaa tgtgcactgc atagacacca ctgtatgcaa ttacagaagc 60
ctgtgagtga acgggggtgt ggtcagtgcg ggcccatggc ctggctgtgc atttaccgaa 120
30 gtctatgagt gaatgggggt gtggtcagtg cgggcccacg gcctggctgg gcctgggagg 180
tttctgatgc tgtgaggcag gagggaagg agggtagggg atagacagtg gtagcccca 240
ccctggaaga cataacagta agtccaggcc cgaaggcag cagggatgct gggggccca 300
cttgggcggc ggggatgatg gagggcctgg ccagggtggc agggatgatg ggggccccag 360
ctgggggtggc aggggtgatg gggggggctg gtctgggtgg cggggaagat ggggaagcct 420
35 ggctgggccc cctcctcccc tgccctccac ctgcagccgt ggatccggat gtgcttccct 480
gggtgcacatc ctctgggcca tcagctttca tggaggtggg gggcaggggc atgacacatc 540
cctgtataaa atccaggatt cctcctctg aacgcccaca ctcaggttga aagtcacatt 600
ccgctctggt ccattctctt aagagtagac caggattctg atctctgaag ggtgggtagg 660
gtggggcagt ggagggtgtg gacacaggag gcttcagggt ggggctgtgt atgctctctc 720
40 atcctcttat catctccag tctcatctct catcctctta tcactccca gtctcatctg 780
tcttctctct atctccagt ctcatctgtc atcctcttac catctcccag tctcatctct 840
tatctcttta tctcctagtc tcattccagac ttacctcca gggcggtgtc caggctcgca 900
gtggagctgg acatacgtcc ttctcaggc agaaggaaat ggaaggattg cagagaacag 960
gagggggcggc tcagagggac gcagtcttgg ggtgaagaaa cagccccctc tcagaagtgt 1020
45 gcttgggcca cagaaaccg agggccctgc gtgagtggct ccagagcctt ccagcaggtc 1080
cctgtgtggg ccttatggta tggccgggtc ctactgagt caccctggac agggcttctg 1140
gtttgagtgc agcccgagc tgccctgtgt cggggtgggg gcttatggcc actggatatg 1200
gcgtcattta ttgtgtctgc ttacagaaat gtctgagtga ccgagcctaa tgtgtatggt 1260
gggcccgaat ccacagactg tgtcgtaaat gcactctggt gcctggagcc cccgtatagg 1320
50 agctgtgagg aaggaggggc tcttggcagc cggcctgggg gcgccttgc cctgcaaac 1380
ggaaggaggc gggcccgggc gccgtgggag gacgacctca agtgagagggt tggacagaac 1440
agggcgggga cttcccaggc gcagaggcgg ctgctcaggc acacctgggt ttgaatcaca 1500
gaccaacagg tcaggccatt gttcagctat ccactctcta caaagctcca gattcctgtt 1560
tctccgggtg ttttttgggt aaattttact caggattact tatatttttt gctaaagtat 1620
55 tagaccctta aaaaaggat ttgtttgat atggcttaac tactaagca cctactttat 1680
ttgtctgttt ttatttatta ttattattat tattagagat ggtgtctact ctgtcacc 1740
gggtgttagt gcagtggcac agtcatggct cgctgtagcc gcaaaccccc aggtcctagt 1800
gatcctccgg cctcagcttc ccagagtgtt gggattacag gtgtgagcca ctgcccctgc 1860
ctggcacttt taaaaaccac tatgtaaggc caggctccag ggcttccaca cctgtcatcc 1920
60 cagtgtttt ggaagccgag gcagaaggat tgtctgaggc caggagtttg agaccagcat 1980
gggtaacata gggagacccc atctctacaa aaaatgcaaa aagtattccg ggcgtgggg 2040
ccagcatctg tagtcccagc tgctcgggag gctgagtggg aggatcgctt gagcccgga 2100
gggtatggct gcagtggact gtgattgtac catcgactc cagcctgggc aacagagtga 2160
gacctgtctt caaaaaaaa aaaaaaaa gaaggagaa gagaaagaa gaagaaggaa 2220
65 gaaggaaa gaagaaga gaagaaggaa gaaagaaggaa gaaggaggcc tgctaggtgc 2280

18 / 18

taggttagact gtcaaatctc agagcaaaat gaaaataaca aagttttaaa gggaaagaaa 2340
aaccaccagct ctttggactt ccttaggcct gaacttcac tcaagcagct tccttccaca 2400
gacaagcgtg tatggagcga gtgagttcaa agcagaaagg gaggagaagc aggcaagggt 2460
ggaggctgtg ggtgacacca gccaggaccc ctgaaaggga gtggttggtt tcctgcctca 2520
5 gcccacgct cctgccggtc ctgcacctgc tgaaccgtc gatggttggtg ccagggtgcc 2580
acctgggaag gatgctgtgc agggggcttg ccaaactttg gtgggtttca gaagcccag 2640
gcacttgtgg caggcacaat tacagccctt ccccaaagat gccacgtcc ttctcctgga 2700
acctgtgaat gtgtcacccg caaggcagag gctggtgaag gctgcagggtg gaatcacggc 2760
tgccagtcag ccgatcttaa ggtcatcctg gattatctgg tgggcctgat atggcccaa 2820
10 gggtccttag aagtgaaga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880
cactggccac tgctggcttt gagatggagg aggggggtccc cagccaagga atgggggag 2940
ccgtccatg ctggaaaagc aagcaatcct ccccggtcct gagggcacac gccctgcc 3000
acgcctcgat ttcaggccag tgggacctgt ttcagcttcc cggcctccag agctgtaaga 3060
tgatgcgttt gtgttcagcc actaagctgc agtgattcgt cacagcagca aatggaatag 3120
15 cagtacaggg aaatgaatac agggacagtt ctcagagtga ctctcagccc acccctggg 3179